HOKKAIDO UNIVERSITY

Title	論文評価指標 ”Knowledge Pool” の課題解決 ： 被引用欠損率との相関および時間補正に関する調査結果
Author(s)	清重, 周太郎
Description	2020年度第1回事業データベース作業部会 口頭発表資料
Issue Date	2020-04-08
Doc URL	https://hdl.handle.net/2115/77339
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	conference presentation
File Information	10_19H00496_paper_v1.1.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

論文評価指標“Knowledge Pool Model”の課題解決
―被引用欠損率との相関および時間補正に関する調査結果－

清重 周太郎†

<抄録>
論文が研究活動に与える影響を表す指標“Knowledge Pool Model”の実用化を阻む要因に関する調査を目的とし，被引用欠損率の変化が Knowledge Pool Model 算出値の信頼性に与える影響を調査した．加えて，発行年が異なる論文間における Knowledge Pool Model 算出値の比較可能性を保持するため，時間補正式の導出を試みた．
<キーワード> 計量書誌学，研究評価，引用ネットワーク，オープン・サイテーション

Problem solving of "Knowledge Pool Model", an index to measure the importance of a paper:
Study of correlation with cited loss ratio, and time correction

Shutaro Kiyoshige

1. はじめに

1.1 背景

インパクトファクターに代表される学術雑誌の影響度指標は長く研究評価を支えてきたが，オープンサイエンスの潮流においては，論文に加え研究データの公開が求められる等，研究成果の形式・公開手段の多様化が進み，雑誌単位に限定されない影響観測の必要性が高まっている（林，2016）．

Knowledge Pool Model（KPM）は 2006 年に角田らによって提案された[1]．これはある論文が他の研究活動に与えた影響を表す指標で，被引用数を元に，引用先の影響度の強弱によって重み付けを行い算出される．重み付けは Google が検索結果の表示順補正に用いる手法 Page Rank を参考としており，引用情報さえ確かならば，掲載誌に左右されず，研究データにも適用できる．2006 年当時，遡及的に電子化された論文の引用情報には不備が散見され KPM の実用化を阻んでいたが，2010 年に国際標準規格となったデジタルオブジェクト識別子（Digital Object Identifier: DOI）の普及によりこの問題は解決に向かっており，オープンサイエンスに対応する KPM の実用化が急務である．実用化を阻む要因は種々あるが，本研究では以下の 2 点について調査を行った．

1.2 被引用欠損率

被引用数 C のうち，被引用情報が適切に整備されているリンク数を L としたとき，リンクが構成されない比率を被引用欠損率（Cited Loss Ratio: CLR）と定める．CLR は次の式で与えられる．

$$CLR = 1 - L/C \qquad (1.1)$$

被引用欠損が起こる背景には，引用情報の誤謬あるいは引用データベースの収録対象外である点が挙げられる．

1.3 時間補正

角田らも指摘する通り，KPM は発行年を遡るほど大きくなる性質が確認されている．発行年が異なる論文同士を比較するためには，発行年の差に基づく補正を行う必要がある．

2. 方法

Web of Science Core Collection（WoSCC）の被引用データを用いて，CLR の変化が KPM に与える影響を調査するとともに，時間補正式の導出を行った．

2.1 検証データ選定

選定条件と結果を**表**に示す．

表 検証データの選定条件

コレクション： Web of Science Core Collection 発行年：1975-2015 年　※論文件数は 2019.8 時点のもの		
分野	キーワード	論文件数
生命科学	Stroke, Rehabilitation	10,107
環境科学	Escherichia, Environment	6,534
経済学	Educational, Business	7,636

2.2 引用ネットワーク整備

検証データの引用情報に基づき被引用ネットワークを整備した．引用情報の誤謬や WoSCC 収録対象外で引用情報が欠ける等，取得データにおいて不備がある場合は適宜修正を加えた．リンク数は生命科学 472,979 件，環境科学 251,529 件，経済学 239,260 件となった．

2.3 比較値の算出と収集

検証データについて CLR を算出し，各分野グループにおける CLR 値の下位 500 件に対して KPM を算出した．KPM の算出には Python の疎行列ライブラリ“scipy.sparse”を用いた[2]．同対象について，Dimensions から相対引用率（Relative Citation Ratio: RCR）[3]を取得した．

2.4 時間補正式

生命科学群において，出版年が 1991-2000 年と 2001-2010 年のグループに分け，KPM 値が近しくなるような時間補正式を導出した．再度 1996-2000 年，2001-2005 年，2006-2010 年の 5 年ごとのグループに対して導出式を適用し，検証を行った．

3. 結果

3.1 CLR と KPM の相関

得られた各数値を比較したところ，被引用数 C と KPM，および RCR については相関が見られた．被引用欠損率 CLR の高低による KPM の信頼性については，有意な差異は認められず仮定と反した．これは，KPM が被引用に重みづけを行う性質から，影響度が高い論文からの引用情報が整備されていれば，その他の被引用情報の有無は KPM 値に大きな影響を与えないことを意味する．

3.2 時間補正式

出版年が A 年である論文の KPM_A を，B 年出版であるものの KPM_B と比較するために補正した値 $KPM_{A'}$は次の式で与えられる：

$$KPM_{A'} = KPM_A \wedge (1 / (1 + ((B-A) * 0.00483))) \quad \text{ただし } (B-A)>0 \qquad (3.1)$$

4. 考察

一般的に，分野への影響度が高いと認められる論文は引用情報の整備が進むことから，あるいは整備が進んだことで発見性が向上しさらなる引用を呼ぶとも考えられるが，いずれにしても情報基盤における発見性保持の価値が大きいことが示唆された．なお，研究データの KPM 算出を試みたが，研究データの引用表示は Reference List に欠けており，分野知識曹 Q へのリンク比率が大きくなったため，意義のある KPM 値は得られなかった．本研究を通じて，引用表示における DOI 等識別子の付与，ならびにオープン・サイテーションの重要性を確認した．

5. 謝辞

本研究は，2019 年度科学研究費補助金 奨励研究（19H00496）の助成を受けている．

6. 参考文献

[1] 角田裕之.「論文と研究者のインパクトに対する新しい計量書誌学的指標」,情報メディア研究,5(1),p1-20,2006.
[2] https://docs.scipy.org/doc/scipy/reference/sparse.html , (参照 2020-3-31)
[3] “What is the RCR? How is the RCR score calculated?”. https://dimensions.freshdesk.com/support/solutions/articles/23000018841-what-is-the-rcr-how-is-the-rcr-score-calculated- , (参照 2020-3-31)
[4] “About the Eigenfactor Project”. http://www.eigenfactor.org/about.php , (参 照 2020-3-31)

本著作の著作権は著者に帰属します．本著作の内容物はクリエイティブ・コモンズ表示 4.0 国際（CC BY 4.0）ライセンスの下に提供されます．
https://creativecommons.org/licenses/by/4.0/deed.ja

† きよしげ しゅうたろう 北海道大学附属図書館利用支援課 北方資料担当（〒060-0808 札幌市北区北 8 条西 5 丁目）
E-Mail: kiyoshige@lib.hokudai.ac.jp
ORCID iD: 0000-0002-9228-6354