Journal of Informetrics 8 (2014) 123–135

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Classification method for detecting coercive self-citation in journals

Tian Yu[a,*], Guang Yu[a,**], Ming-Yang Wang[b,1]

[a] School of Management, Harbin Institute of Technology, 92 West Dazhi Street, Nan Gang District, Harbin 150001, PR China
[b] College of Information and Computer Engineering, Northeast Forestry University, No. 26 Hexing Road, Xiangfang District, Harbin 150040, PR China

ARTICLE INFO

Article history:
Received 24 August 2013
Received in revised form 29 October 2013
Accepted 4 November 2013
Available online 1 December 2013

Keywords:
Scientific journal
Coercive self-citation
Classification

ABSTRACT

Journal self-citations strongly affect journal evaluation indicators (such as impact factors) at the meso- and micro-levels, and therefore they are often increased artificially to inflate the evaluation indicators in journal evaluation systems. This coercive self-citation is a form of scientific misconduct that severely undermines the objective authenticity of these indicators. In this study, we developed the feature space for describing journal citation behavior and conducted feature selection by combining GA-Wrapper with ReliefF. We also constructed a journal classification model using the logistic regression method to identify normal and abnormal journals. We evaluated the performance of the classification model using journals in three subject areas (BIOLOGY, MATHEMATICS and CHEMISTRY, APPLIED) during 2002–2011 as the test samples and good results were achieved in our experiments. Thus, we developed an effective method for the accurate identification of coercive self-citations.

1. Introduction

Scientific journals are media used to report academic achievements so they play very important roles in scientific development. Journal self-citations are citations of previous papers in the same journal, which are used to highlight research results and to stimulate a specific research direction. Thus, journal self-citation reflects the progress of scientific research and it satisfies the rules of scientific development. In academic circles, the impact factor is a commonly used scientific index to evaluate journals, which also considers the quantity of self-citations. However, certain individual journal editors have disregarded the moral principles of scientific integrity and forced authors to add one or more articles from their own journals as recently published references (Arnold, 2009). The coercive self-citations may call into question the citation evaluation system that scientists have used for many years. Journals that engage in manipulation by self-citation can make the performance look better than the actual impact factor, while some editors and publishers force authors to cite articles published in their journals for commercial purposes (Opthof, 2013).

Science has its own rules and journals are respected based on their citation performance. However, coercive self-citations contradict the normal rules of scientific progress and seriously undermine the integrity and impartiality of journals. These artificial citations cause journals to lose their original scientific value and produce a detrimental academic atmosphere. Thus,

* Corresponding author. Tel.: +86 0451 86412498.
** Corresponding author. Tel.: +86 0451 86413537.
E-mail addresses: yutian.hit@gmail.com (T. Yu), yug@hit.edu.cn (G. Yu), wangmingyang@nefu.edu.cn (M.-Y. Wang).
[1] Tel.: +86 13100966964.

1751-1577/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.11.001

it is necessary to detect and monitor coercive self-citations. In recent years, researchers have paid increasing attention to self-citation, particularly the causes and impacts of self-citation. To the best of our knowledge, very few studies have addressed the identification of abnormal journals involved with coercive self-citation. The purpose of the present study was to detect, expose, and stop this undesirable behavior, thereby ensuring that scientific journals present their actual performance in an honest manner to facilitate normal scientific progress.

The aim of this study was to develop a method for automatically identifying journals that engage in coercive self-citation using a classification model, which was based on information related to known journals. An exploratory analysis using classification techniques was performed to identify normal and abnormal journals. Next, we reviewed studies related to journal self-citation. We then established a feature space to describe journal citation behavior using the actual data for journals indexed in JCR. We presented the methods used to develop the classification model, which identified the key features required to establish the journal classification model. Finally, we provided our conclusions.

2. Literature review

2.1. Journal self-citations and journal impact factor

Previous studies of journal self-citation have focused mainly on the rules of self-citation behavior. Traditionally and logically, it has been assumed that self-citation is permissible and necessary. The normal use of self-citations facilitates knowledge flow and diffusion, which are useful relevant and functional. Hyland (2003) suggested that self-citation is a significant means of promoting one's scholarly academic reputation and gaining professional credit for one's research. Carley, Porter, and Youtie (2013) stated that the effect of self-citation in the measurement of research production was very important because the effects of self-citation could be very significant at the meso- and micro-levels. Self-citation rate statistics and comparisons of different subjects across years have shown that the proportion of self-citations ranges from 7% to 20% and they vary significantly among different scientific disciplines. Thus, it is necessary to use self-citation rates carefully during evaluations (Aksnes, 2003; Garfield, 1996, 1997; Glänzel, 2004; Tiew, 2000).

In recent years, journal evaluation systems based on quantitative analysis have played increasingly important roles in journal evaluation. Thus, the relationships between journal self-citations and journal evaluation indicators have been determined over various time periods, particularly for the journal impact factor. There is still no final conclusion on how journal self-citations affect the impact factor because researchers have obtained variable results in different research fields. Thus, understanding how journal self-citation affects journal impact factor demands more research. In some subject areas, researchers have shown that self-citation is significantly correlated with journal impact factor and that the value of the impact factor changes markedly with inclusion or exclusion of self-citations (Frandsen, 2007; Kurmis & Kurmis, 2010; Mehrad & Goltaji, 2010; Straub & Anderson, 2009). In other subject areas, however, researchers have suggested that self-citations have insignificant effects (Brice & Bligh, 2004; Fan & McGhee, 2008; Finardi, 2013; Frandsen, 2007; McVeigh, 2004; Motamed, Mehta, Basavaraj, & Fuad, 2002) or only weak effects on the journal impact factor (Campanario, 2010).

However, only normal self-citations were taken into account in the studies mentioned above. In fact, not all journal self-citations are normal and necessary, and abnormal journal self-citations can strongly affect the journal impact factor (Falagas & Alexiou, 2008; Opthof, 2013; Smith, 1997).

2.2. Journal coercive self-citation

Journal coercive self-citation is a kind of abnormal citation. The publishers and editors of some journals usually implicitly or directly require that a manuscript cite at least one paper published in their own journal to increase the cumulative citation count, or else the paper will be rejected. This is called journal coercive self-citation behavior (Smith, 1997). Obviously, the main purpose of coercive self-citations is to make the journal impact factor increase in a short time (Falagas & Alexiou, 2008; Opthof, 2013).

In 1955, the journal impact factor was proposed by Eugene Garfield. His original intention was to develop an index to measure the quality of scientific journals and to select journals for the Science Citation Index (Garfield, 2006). Despite its shortcomings, the impact factor soon became widely accepted as the gold standard of quality scientific publishing and of a journal's prestige (Campbell, 2005; Romano, 2009). However, when the journal impact is inappropriately attributed to all articles published in a single journal, it leads to false applications regarding the evaluation of individual researchers or research groups. Unfortunately, this is a common practice, especially when government funding committees and academic institutions judge and fund. As the use of an evaluation index has become excessive, the impact factor has accumulated huge power and importance (Arnold, 2009; Falagas & Alexiou, 2008). Therefore, the business interests behind the impact factor are the root cause of the occurrence of journal coercive self-citation.

Due to design flaws and deficiencies in its defining equation, the impact factor can be inflated rapidly in a short time by manipulation, which gives a journal a false "better" performance value in the journal evaluation system (Sevinc, 2004). It has been found that a considerable portion of scientific journals are involved in the "impact factor game" (Chang & Maasoumi, 2013). Wilhite and Fong (2012) surveyed researchers from disciplines such as economics, sociology, psychology, and business to explore the extent and nature of coercive self-citations. Their results showed that this type of behavior was widespread and over 20% of all journals were involved. Less than 7% of the respondents refused to add superfluous citations when asked