Journal of Informetrics 9 (2015) 449–454

Contents lists available at ScienceDirect

Journal of Informetrics

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

Compressing multiple scales of impact detection by Reference Publication Year Spectroscopy

Jordan A. Comins*, Thomas W. Hussey

Virginia Tech Applied Research Corporation, Arlington, VA, United States

ARTICLE INFO

Article history:
Received 20 January 2015
Received in revised form 24 February 2015
Accepted 20 March 2015

Keywords:
Scientometrics
Informetrics
Citation analysis
Reference Publication Year Spectroscopy
Science policy
Funding agencies

ABSTRACT

Reference Publication Year Spectroscopy (RPYS) is a scientometric technique that effectively reveals punctuated peaks of historical scientific impact on a specified research field or technology. In many cases, a seminal discovery serves as the driving force underlying any given peak. Importantly, the results from a RPYS analyses are represented on their own distinct scales, the bounds of which vary considerably across analyses. This makes comparing years of punctuated impact across multiple RPYS analyses problematic. In this paper, we propose a data transformation and visualization technique that resolves this challenge. Specifically, using a rank-order normalization procedure, we compress the results of multiple RPYS analyses into a single, consistent rank scale that clearly highlights years of punctuated impact *across* RPYS analyses. We suggest that rank transformation increases the effectiveness of this scientometric technique to reveal the scope of historical impact of seminal works by allowing researchers to simultaneously consider results from multiple RPYS analyses.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Citations signify the relevance of prior research or invention. In the scientific community, the aggregation of citations attributed to a specific work is commonly taken as a central indicator of its scholarly impact (De Solla Price, 1965; Garfield, Malin, & Small, 1978; Radicchi, Fortuno, & Castellano, 2008). More generally, citations and citation counts are thought to represent how knowledge accumulates, combines and transfers to generate new ideas and discoveries. Since citations function as linkages between scientific works, citation records provide an opportunity to quantitatively identify seminal contributions to a given research field or technology (Kostoff & Shlesinger, 2005; Marx, Bornmann, Barth, & Leydesdorff, 2014; van Raan, 2000).

One technique leveraging citations to detect important scientific contributions is Reference Publication Year Spectroscopy (RPYS). RPYS offers a quantitative approach to assist in identifying the historical roots of research fields and topics (Marx et al., 2014). To accomplish this, RPYS considers the references cited by a cohort of publications resulting from a particular database query. By way of example, consider a search query for topic X that returns only one article. If this article, published in 1990, cites a reference published in 1980, then 1980 is used as a data point in the foregoing analysis. As such, after a set of publications is returned from a database query, the publication date of each cited reference from this set of publications is extracted and mapped onto a frequency distribution sorted by time. The resulting visualization often reveals punctuated

* Corresponding author at: Virginia Tech Applied Research Corporation, Arlington, VA 22203, United States. Tel.: +1 703 879 8142.
E-mail address: jcomins@gmail.com (J.A. Comins).

http://dx.doi.org/10.1016/j.joi.2015.03.003
1751-1577/© 2015 Elsevier Ltd. All rights reserved.

Table 1
Search terms and the corresponding number of results yielded from the Web of Science Core Collection (1974-present).

WoS core collection topic search	No. of results (total)	No. of results (articles and conference proceedings only)	No. of cited references obtained
"Viterbi"	4840	4713	62,405
"convolutional code" OR "convolutional codes"	4011	3863	59,586
"hiddenmarkov model" OR "hidden markov models"	2348	2312	58,648
"continuous speech recognition"	1569	1538	25,070
"automatic speech recognition"	3427	3370	66,233
"speech recognition"	17,946	17,462	330,250

peaks in the distribution. These peaks correspond to years containing a larger number of cited references within discrete bins of time. Often, these peaks are driven by a large number of references to a seminal work in the field. To date, RPYS has been successfully applied to investigations of important early contributions in several research topics (Leydesdorff, Bornmann, Marx, & Milojević, 2014; Marx et al., 2014; Marx & Bornmann, 2013; Wray & Bornmann, 2014).

There is, however, a major challenge with the current methodology. Namely, the results produced by a given RPYS analysis are represented within their own distinct range or scale, the bounds of which vary considerably across analyses. In other words, using the presently defined RPYS technique, it could be difficult to compare patterns of maxima for the cited references of a small research fields with those of the cited references from a much larger research fields. Making RPYS analyses amenable to large-scale comparative analysis is an important extension of the technique for future applications. For instance, it would allow analysts to more readily evaluate whether a large number of research topics show a similar history of important findings as demonstrated through cited works. A second possibility is that being able to estimate the similarities between the citation histories for various subfields might open-up an entirely novel venue for defining the relationships between these subfields – with the assumption being that research areas correspondingly informed by the same seminal works are more similar than those that are not.

To address this shortcoming, we demonstrate here how the addition of a simple data transformation procedure to the standard RPYS methodology can aid in the detection of shared patterns of maxima for the cited references across RPYS analyses, which potentially suggest common historical influences. Specifically, we adopt the use of a rank-transformation procedure commonly used in inferential statistics to perform non-parametric analyses (Conover & Iman, 1981; Labovitz, 1970). This transformation compresses the multiple scales produced from various RPYS analyses into a single rank scale that allows researchers to identify years of punctuated impact *across* RPYS analyses. We describe a visualization procedure that efficiently represents data from multiple RPYS analyses concurrently.

To demonstrate the efficacy of this procedure, we begin with a publication that we suspect a priori has meaningfully impacted numerous research topics: the Viterbi algorithm first published by Andrew Viterbi in 1967. In this groundbreaking work, Viterbi describes an algorithm that identifies the most likely sequence of hidden states associated with a sequence of known or observed states. The algorithm is widely used in stochastic models and error-correcting (or decoding convolutional codes) as well as in a variety of computational procedures pertaining to machine speech recognition (Viterbi, 2006). Given this, we performed six RPYS analyses for research topics pulled from the Web of Science that all pertained to the development or use of the Viterbi algorithm (Viterbi, 1967). The Viterbi algorithm's impact and use in a wide array of research communities, from statisticians studying stochastic models to engineers and computer scientists working on various aspects of machine speech recognition, make it an ideal candidate for demonstrating the value of this data-transformation procedure for comparing multiple RPYS analyses concurrently.

2. Method

We accessed and downloaded data from the Thomson Reuters Web of Science (WoS) between December 12, 2014 and December 14, 2014. We performed topic searches using the Web of Science Core Collection, for which we had back-records from 1974 to 2014. The topic searches performed for our six RPYS analyses were as follows: (1) "Viterbi", (2) "convolutional code" OR "convolutional codes", (3) "hiddenmarkov model" OR "hidden markov models", (4) "continuous speech recognition", (5) "automatic speech recognition" and, finally, (6) "speech recognition". The number of results for these six searches is shown in Table 1. We then filtered these results to only represent articles and conference proceedings (again, values in Table 1). The last step of the data collection process was to download the results from WoS Core Collection using the "Save to Other File Formats" option and selecting "Full Record and Cited References" as our desired record content.

To extract cited references from the data we obtained, we used the open source Sci2 tool (Sci2 Team, 2009) developed by Indiana University. This converts the Full ISI records into tables. This conversion conveniently represents "Cited References" as a single column within this table. The resulting data contained in this column uses a | character as a delimiter between the different cited references. Using a python script, we split our cited references data using this delimiter to generate a new table representing all of the cited references. This process allows us to effectively isolate properly structured references and their associated publication year from WoS records.

Download English Version:

https://daneshyari.com/en/article/10358521

Download Persian Version:

https://daneshyari.com/article/10358521

Daneshyari.com