Original Investigation

Use of a Machine-learning Method for Predicting Highly Cited Articles Within General Radiology Journals

Andrew B. Rosenkrantz, MD, MPA, Ankur M. Doshi, MD, Luke A. Ginocchio, Yindalon Aphinyanaphongs, MD, PhD

Rationale and Objectives: This study aimed to assess the performance of a text classification machine-learning model in predicting highly cited articles within the recent radiological literature and to identify the model's most influential article features.

Materials and Methods: We downloaded from PubMed the title, abstract, and medical subject heading terms for 10,065 articles published in 25 general radiology journals in 2012 and 2013. Three machine-learning models were applied to predict the top 10% of included articles in terms of the number of citations to the article in 2014 (reflecting the 2-year time window in conventional impact factor calculations). The model having the highest area under the curve was selected to derive a list of article features (words) predicting high citation volume, which was iteratively reduced to identify the smallest possible core feature list maintaining predictive power. Overall themes were qualitatively assigned to the core features.

Results: The regularized logistic regression (Bayesian binary regression) model had highest performance, achieving an area under the curve of 0.814 in predicting articles in the top 10% of citation volume. We reduced the initial 14,083 features to 210 features that maintain predictivity. These features corresponded with topics relating to various imaging techniques (eg, diffusion-weighted magnetic resonance imaging, hyperpolarized magnetic resonance imaging, dual-energy computed tomography, computed tomography reconstruction algorithms, tomosynthesis, elastography, and computer-aided diagnosis), particular pathologies (prostate cancer; thyroid nodules; hepatic adenoma, hepatocellular carcinoma, non-alcoholic fatty liver disease), and other topics (radiation dose, electroporation, education, general oncology, gadolinium, statistics).

Conclusions: Machine learning can be successfully applied to create specific feature-based models for predicting articles likely to achieve high influence within the radiological literature.

Key Words: Radiology; bibliometrics; biomedical journals; machine learning.

© 2016 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

INTRODUCTION

Identification of articles having the greatest impact, as measured through future citations, has been a topic of interest within the radiological literature. Three recent studies have explored the most highly cited radiological articles from a broad historical perspective, evaluating a large number of journals over an extended time frame (1–3). At least three additional articles have explored the most highly cited radiological articles within a specific journal of subspecialty area (4–6). Such investigations seek to provide insights that will be helpful for researchers in shaping their studies, for journal editors and reviewers in selecting high-impact journal content, and for radiologists in appreciating the topics of greatest interest in the field (1,3). However, all of these earlier studies used essentially the same approach in their analyses: manually evaluating the content of solely the 100 most highly cited articles relevant to the question at hand. Although providing useful information from a historical perspective, this approach draws conclusions regarding the imaging literature based on only an extremely small fraction of available published articles. Moreover, a simple manual coding of features of the most highly cited articles, although easy to perform, risks leading to incorrect information regarding those features that in fact have the greatest influence in predicting a high-citation volume. In addition, the manual approach does not provide any numerical estimate of the actual overall performance of the identified features in predicting citation volume. Finally, the approach fails to provide any quantitative method for evaluating the likelihood of a given article to be highly cited. Such ability would be valuable for investigators aiming to maximize the impact of their research and for editors and reviewers aiming to enhance a journal's status.

Machine learning provides an alternate approach for performing a sophisticated evaluation of citation frequency based on a broad spectrum of article features. This approach

Acad Radiol 2016; ■:■■–■■

From the Department of Radiology, NYU Langone Medical Center, 660 First Avenue, 3rd Floor, New York, NY 10016 (A.B.R., A.M.D., L.A.G.); Center for Healthcare Innovation and Delivery Science, NYU Langone Medical Center, New York, New York (Y.A.). Received June 25, 2016; revised August 18, 2016; accepted August 19, 2016. **Address correspondence to:** A.B.R. e-mail: Andrew.Rosenkrantz@nyumc.org

© 2016 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.acra.2016.08.011

generates computational models for predicting citation volume based on text analysis of the article's title and abstract in combination with other article meta-data, using a very large number of articles (7). This comprehensive scheme can identify trends that would be difficult for human observers to otherwise detect. Perhaps more important in comparison to the historical approaches of the earlier cited radiological studies, the actual predictive performance of a machine-learning model can be quantified, and the model can be applied prospectively to predict future citations at the time of an article's publication, if not even earlier (7). As an example, a machine-learning tool to automatically predict citations could be used to efficiently and reliably aid the initial triage and assessment of submitted manuscripts.

Past works have demonstrated the feasibility of advanced computer algorithms in bibliometric inquiries in other biomedical disciplines, consistently achieving a favorable predictive performance (7–9). Application of these methods within radiology may provide further insights to complement those of the earlier radiological bibliometric investigations that relied on manual assessment of a small number of articles. Therefore, in this study, our aim was to assess the performance of a text classification machine-learning model to predict highly cited articles within the recent radiological literature and to identify the model's most influential article features.

METHODS

Corpus Construction

This retrospective study did not require institutional review board approval, as it did not involve human subjects research. The 25 general radiology journals having the highest impact factor (IF) in 2014 (Table 1) were selected for inclusion based on their assigned category of "Radiological and Nuclear Medicine Sciences" within Thomson Reuters' Journal Citation Reports (10). Journals within this category not related to clinical radiology (eg, radiation oncology, molecular imaging, or optics) were excluded. In addition, organ- or modality-based subspecialty radiology journals, which tend to have different citation patterns (11,12), were not included (aside from the journal *Diagnostic & Interventional Radiology*, which includes both general diagnostic and interventional radiology content, comparable to a number of other included general radiology journals). Finally, a general radiology journal, the journal *Current Medical Imaging Reviews*, was excluded owing to its articles not being indexed on PubMed, which served as the data source for key elements of the study, as explained below.

Web of Science was used to identify all articles published in included journals in 2012 and 2013, as well as the number of citations to these articles in 2014 (13). A total of 11,318

TABLE 1. List of Included Radiological Journals and Associated 2014 Impact Factors, Ranked in Order of Descending Number of Articles in the Top 10% of All Articles in the Analysis

Journal	20014 Impact Factor*	Number of Included Articles From 2012 or 2013	Number of Included Articles in Top 10% in Terms of Citations to Articles in 2014
Radiology	6.867	998	387
European Radiology	4.014	746	200
American Journal of Roentgenology	2.731	1320	178
European Journal of Radiology	2.369	1363	178
Investigative Radiology	4.437	220	75
British Journal of Radiology	1.984	575	65
RadioGraphics	2.602	312	45
Clinical Radiology	1.759	506	42
Academic Radiology	1.751	469	36
Journal of the American College of Radiology	2.836	518	31
Radiologic Clinics of North America	1.984	139	21
Acta Radiologica	1.603	382	19
Korean Journal of Radiology	1.571	292	13
Diagnostic and Interventional Radiology	1.436	178	11
Japanese Journal of Radiology	0.837	255	9
Radiologia Medica	1.343	221	8
Rofo	1.402	283	8
Clinical Imaging	0.810	379	6
BMC Medical Imaging	1.312	80	3
Surgical and Radiologic Anatomy	1.047	289	3
Seminars in Roentgenology	0.705	81	1
Canadian Association of Radiologists Journal	0.519	129	0
Iranian Journal of Radiology	0.366	86	0
Radiologe	0.425	244	0

* Identified from reference 10.

Download English Version:

https://daneshyari.com/en/article/5725773

Download Persian Version:

https://daneshyari.com/article/5725773

Daneshyari.com