The 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013)

A Machine Learning Based Method for Optimal Journal Classification

Saeed Iqbal, Muhammad Shaheen, Fazl-e-basit
(National University of Computer and Emerging Scinece) Department of Compter Science
FAST-NUCES
Peshawar, Pakistan
saeediqbalkhattak@gmail.com, muhammad.shaheen@nu.edu.pk, fazl.basit@nu.edu.pk

***Abstract*—We present a hypothetical and realistic examination and exploration of a number of bibliometric indicators of journal performance. In this paper, the indicators we have focused upon are Eigenfactor indicator, Impact factor, audience factor and Article influence weight indicator. Our focus is to find the missing parameters and some limitations that have not been conducted in previous algorithms. To find the influential parameters and to propose a new journal performance factor, that ranked a journal in best accepted manner. For calssification and verification purpose we use a machine learning classification technique (Bayesian classification). It is one of the most common learning algorithms in machine learning classification. Using bayesain classification, we classify several journals according to our proposed methods and compare results with the previous methods.**

***Index Terms*—Journal ranking, Impact Factor, Eigenfactor, Article Influence, Prestige of Journal**

I. Introduction

In research evaluation system, citation analyses play a vital role. For professional societies, scholarly institutions, individual scientists, academic department and for funding sources qualitative research is important. For qualitative and quantitative analyses of journals there are some databases that maintain data about articles, authors, citations, journals, countries, etc.

The most commonly used database is Journal Citation Report (JCR) [1] and the Science Citation Index (SCI) [2] produced by the Thompson Scientific(formaly Institute for Scientific Information (ISI)) [3].

The most extensively used and influential scientometric indicator is journal impact factor (IF) [4]. According to the following equation the Impact Factor is calculated every year. As mentioned in equation 1

$$IF(Y) = \frac{\text{Citations in Y to documents published in } Y_1 \text{and } Y_2}{\text{Citable items published in } Y_1 \text{and } Y_2} \quad (1)$$

Where Y_1 and Y_2 are the two years before Y. For example to calculate impact factor of journalX in 2012. As mentioned in equation 2

$$IF(jX-2012) = \frac{\text{Tot. Cit. in 2012 of jX art. pub. (2010-11)}}{\text{Tot. citbl. jX art. that are pub. (2010-11)}} \quad (2)$$

Thompson Scientific (formerly ISI) includes citable items, articles, notes, and reviews to calculate the denominator, and count all citations received by the journal to calculate the nominator. The ratio between citations and recent citable items published by a journal is the impact factor. The average article in a journal has been cited in a particular period is measured by the impact factor.

In the Scopus [5] database that contains information about the journal and country scientific communities, the SCImago Journal & Country Rank is a portal that analyzes the journal information and country scientific indicators. This platform is called SCImago Journal Rank (SJR) [6]. In SJR, the details about journal ranking is deliberated in respective manner and is calculated by averaging number of weighted citations received in the selected year by the documents published in the selected journal in last three years.i.e. weighted citations received in year X to documents published in the journal in years (X-1), (X-2) and (X-3).

On the basis of above, journals are classified into several categories. Different data mining techniques are used for several types of categorization and classification. Data mining is a computational process and has analysis steps that extracts patterns, classes and cluster from a large set of data [7], [8], [9]. We analyze and extract patterns, classes and meaningful information from raw and dis-sorted data, then after analysis and application of several other analytical processing techniques the implicit patterns becomes knowledge. When data mining techniques are applied on databases, unknown, non-trivial and potentially useful information is extracted from the database [10], [11], [12].

This paper is organised as follows. Section 2 reviews the background of journal evaulation algorithms and data mining techniques. Section 3 decribes the main theme of this paper in the form of problem specification and description. In section 4, we proposed a method for journal evaulation. Section 5 decribes the implementation of our proposed algorithm. Finally, in section 6, we compared the results of our method with some previous methods, in section 7 we conclude our research work.

II. Background

Around the Impact Factor of journals, there has been substantial discussion and criticism specially to assessing the research output of groups or individuals, it does not deliberate all the key factors that are required for multifaceted task of