Translating Peer Review Reports into Measurable Scores and their Relationship with Citations-Deploying Sentiment Analysis

P Pichappan
Digital Information Research Labs
Chennai
India
{pichappan@dirf.org}

ABSTRACT: *In Scientific evaluation, citation-based indicators are widely deployed across scientific disciplines. They suffer from many limitations and they are extensively documented. Peer review of scientific papers serve as a viable alternative; however, using it for comparative evaluation is a difficult exercise as they do not permit generating numbers or scores which are required for any assessment. In this work, we generated a metric to bring some conversion of text-based review using sentiment analysis and we did a comparison with citation scores. The results and the correlation between these two measures are presented.*

Keywords: Peer Review Scores, Citation Scores, Scientific Paper Evaluation, Sentiment Analysis

DOI: https://doi.org/10.6025/stm/2020/2/91-95

Prelude

Since the evolution of Impact Factor in 1955, many indicators have been proposed out of which the citation based indicator is primarily used. All the proposed indicators have their own value and limitations. Among them the most common indicators are the Publications, Citation-based ones, Patents, Altmetrics, Experts Evaluation and a few more. However, no single indicator is accepted as a precise one.

Peer review in Evaluation

When the experts view or assess a paper, it is considered as most significant. Citations are given to papers because they are relevant to a current paper and no way indicate the scientific value. However, the peer reviews are produced based on the evaluation of scientific content. There are shortcomings in peer review systems also as all reviewers do not have expertise and the review scales are inconsistent.

Primary Questions

The primary questions now we can raise are the following.
Is it possible to convert the text-based system into number-based ones?
If the numbers represent the scientific papers content, how these are reliable?
What is the relation between review scores and citation scores?

Nature of Peer reviews

Peer review reports are normally textual and reviewers do not provide any scores for the papers. Citations are expressed in numbers and journals and authors are ranked based on numbers. However, peer review reports are not expressed in numbers. Most of the reviews are textual and recommendations are given finally to accept or revise or reject submissions.

Some review systems provide the reviewers to assign a rating in numbers.

Openconf asks the reviewers to give final rating in a six criteria.

Reject: 0
Probable Reject: 1
Marginal Tend to Reject:2
Marginal Tend to Accept: 3
Clear Accept: 4
Must Accept: 5

Edas offers the rating in three-point scale
Good-3, Fair-2, Poor -1

Body of the Review

Introduction about the paper- Neutral

Error identification- Negative scores

Originality-Scores possible

Method-Scores possible

Experimentation- Scores possible

Overall contribution- Scores possible

Comments to Authors (mostly suggestions)- Scores not possible

Comments to the editors- Scores possible

Sentiment Analysis

In text processing, several techniques as entity analysis, Syntactic analysis, Content classification, sentiment analysis are used.Sentiment analysis is the most suitable method for review text processing. Sentiment analysis inspects the given text and identifies the views and opinions within the text, especially to determine a respondent's attitude as positive, negative, or neutral. Sentiment analysis is performed through the analyze Sentiment method. Sentiment analysis attempts to determine the overall attitude (positive or negative) expressed within the text. Sentiment is represented by numerical score and magnitude values.We classify the peer reviews as the sentiments of reviewers.

Parameters for Sentiment Analysis

- Sentiment Scores range between -1.0 (negative) and 1.0 (positive) and corresponds to the overall views of the reviewers of the papers. Sentiment scores are normalized.

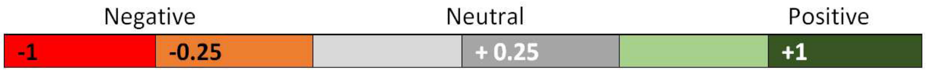
- Magnitude indicates the overall strength of the reviews in terms of length of reviews. The reviews both positive and negative in a given review stands between 0.0 and + “. Unlike score, magnitude is not normalized; each expression of emotion within the text (both positive and negative) contributes to the text's magnitude (so longer reviews likely to have greater magnitudes).

Data Source

We took the reviews of the papers submitted to the International Conference on Digital Information Management (ICDIM) for the period from 2010 to 2015 as the base. We searched in google scholar for the citation record of the papers of ICDIM. We took the top 100 papers and arranged all in the decreasing frequency of citations. The highly cited 26 papers are arranged in decreasing citation order.

The reviews of the 26 papers are extracted from the ICDIM paper submission system. These 26 papers have a mean of 3.15 reviews per paper with a total of 82 reviews. We then subjected these 82 reviews for text analysis using sentiment scores.

Score Range



A review with a neutral score (around 0.0) may indicate a low-review content, or may indicate mixed reviews, with both high positive and negative values which cancel each out. Generally, we can use magnitude values to disambiguate these cases, as truly neutral reviews will have a low magnitude value, while mixed reviews will have higher magnitude values.

Cloud Application

It is possible to prepare a customized NLP application using a large scale text analysis. In such cases all possible expressive words will serve as inputs. These NLP applications use text lexicons such as wordnet, lexical nets etc and assign values that range from 0 to 1. Besides phrases with two or more words will be assigned scores. Since these exercises involve large scale computing programs, we confined our analysis with the use of a pre-determined tool text-to-data which works in cloud.

How the cloud application works?

Each review is subjected to the sentiment analysis.

We use two parameters, Sentiment Score and Magnitude.

Illustration – Case 1

The paper is highly acceptable for originality and novelty The method and implementation are incremental. But the quality of figures is not good. It is reasonably acceptable In general as it is a good paper, and well-structured. I suggest to improve the quality of images and add more newer references. It can be accepted

positive (+0.64) ***Magnitude: 5.86***

Detected Themes	Magnitude	Sentiment Score
Well-structured	1.00	+0.936
Good paper	1.00	+0.748
Reasonably acceptable	0.99	+0.744
Originality and novelty	0.53	+0.506

Detected Keywords	Magnitude	Sentiment Score
Acceptable	0.989	+0.745
Good	0.988	+0.743
Suggest	0.001	+0.250
Structured	0.002	+0.250
More	0.002	+0.249
Improve	0.151	+0.212

Core sentences	Magnitude	Sentiment Score
The paper is highly acceptable for originality and novelty. The method and implementation are incremental	0.69	0.079

But the quality of figures is not good.	0.67	**-0.530**
It is reasonably acceptable In general it is a good paper well structured.	0.93	+0.512
I suggest to improve the quality of images and add more newer references.	0.27	0.183
It can be accepted	0.68	+0.574

Illustration 2 (A rejected paper)

REVIEW: They're may be a need to improve health care web sites, but the paper does not convince me on this. The statistics you present is on a very general level and does not prove much. You say you based you're work to the main search engines, but did only a comparison of the number of hits between different search engines. Instead, you present (more interesting data) data from other sources. You should clarify what you aim to do with this research, and than implement you're study in a robust way (for example by sampling health web sites and analyzing those based on you're criteria.

This document is: **negative (-0.63)**   *Magnitude:* ***2.60***

Core sentences	**Magnitude**	**Sentiment Score**
Their may be a need to improve health care web sites. but the paper does not convince me on this.	0.85	-0.530
The statistics you present is on a very general level. and does not prove much	0.61	-0.471

Results

Cites	**Title**	**Year**	**Sentiment score**	**Magnitude**
173	Wolf search algorithm	2012	+ 0.92	2.73
96	Integrating nature-inspired	2012	+ 0.84	3.82
81	Social network collaborative	2010	+ 0.93	4.98
62	Telecommunication subscribers'	2013	+ 0.91	4.17
53	Elephant search algorithm for	2015	+ 0.88	5.28
42	Muithu: Smaller footprint, ...	2012	+ 0.69	4.09
40	Evaluating recommender....	2012	+ 0.87	6.22
40	Graph mining: A survey of	2012	+ 0.74	6.28
37	Clustering medical data to	2010	+ 0.85	5.86
36	An enterprise business	2010	+ 0.91	6.95
33	Clustering approaches for data	2010	+ 0.88	5.38
32	Opinion mining over twitterspace:...	2012	+ 0.64	0.96
32	Representing history in graph- .	2013	+ 0.61	4.17
31	Network traffic classification	2015	+ 0.58	3.45
31	OpenLabyrinth: An abstract	2010	+ 0.56	2.18

28	A conceptual model for automated .	2014	+ 0.57	1.2
27	Complementarity of process- ...	2010	+ 0.71	6.12
26	Measuring the credibility of	2010	+ 0.65	3.36
25	Real-time business intelligence	2010	+ 0.55	1.88
24	Achieving privacy and security	2012	+ 0.45	1.24
23	A novel feature selection by	2014	+ 0.78	2.19
23	Flatten hierarchies for large-scale . .	2010	+ 0.84	1.02
23	Using Chinese part-of-speech	2010	+ 0.62	3.87
22	HOG and LBP: Towards a robust ...	2015	+ 0.52	2.65
22	Factors influencing	2010	+ 0.43	4.3
20	X-STROWL: A generalized	2012	+ 0.67	5.14

To identify the relevance between citations and peer review scores, we compute correlation. We found that overall correlation is +0.461.

Thus, correlation stands between Moderate and High

Shortcomings

Review reports are available only for the pre-revised submissions, not for revised versions whereas the revised papers are only cited.In the current work, the results are not free from deviation
Error Detection and accuracy rating is not carried out.

Summary and Future Direction

We plan to refine the methods for review score calculation. Review metrics may score over Altmetrics such as downloads, social media, google impact and others. There is a relation between citation and peer review scores. Detailed analysis involving large number of reviews may lead to conclude that review metrics has edge over citation-based indicators. There is a room for peer review scores computation.