Adjusted Impact Factor by using F min Search Algorithm

Woojin Lee, Junghoon Hah
Department of Industrial Engineering, Seoul National University, Republic of Korea

Namhyoung Kim*
Department of Applied Statistics, Gachon University, Republic of Korea

ABSTRACT

Assessing research quality is a crucial task for research institutions. Impact Factor has been widely used as a journal evaluation metric. It also has been used as an evaluation metric for researcher's research ability. However, assessing research quality using Impact Factor can raises various problems, especially when evaluating researchers from different research fields at the same time. For instance, lower Impact Factors are given to traditional engineering and social sciences than those given to general sciences and biology. To solve these problems, adjusted Impact Factor has been proposed. They have the effect of normalizing Impact Factors in different subject categories. However, these revised Impact Factors cause errors in relative rankings among journals. By using linearly combined 11 revised Impact Factor, we find the optimal combination to reduce the error rate and give a solution to weak point of current Impact Factor.

Keywords: Impact Factor, F Min Search, Journal Evaluation Metric

* Corresponding Author, E-mail: nhkim@gachon.ac.kr

1. INTRODUCTION

Assessing research performance is one of the important tasks of research organizations. Many institutions, including universities and public research institutes have developed ways to quantify the quality of research in order to evaluate individual research projects or researchers.

Evaluating individual research performance is a crucial component of research assessment, and outcomes of such evaluations can play a key role in establishing institutional research strategies, including funding plan, hiring, firing, and promotions. However, evaluating individual research performance is a complex task that could not be achieved with a single indicator. There are few consensus and standards accepted simultaneously in various fields to objectively measure the quality of research. In the past, quantity oriented measures have been used as evaluation indicators. For example, the number of papers published

over a period of time was used as an index of evaluation in general. Recently, evaluation methods considering both quantity and quality have been proposed (Sahel, 2011). These methods combine number of publications, number of citations, influence indicators and other indicators together to present quantitative indicator. The number of citations of individual papers, however, is affected by how long the papers have been published. Thus Impact Factor has been widely used as a journal evaluation metric for its efficiency. Impact Factor is defined as the average number of citations of each journal in recent two years to the articles published in that journal. Evaluation of researcher's research ability is assessed by the influence of the Impact Factor score.

This Impact Factor data is provided by the database of Journal Citation Reports from Thomson Reuters. How competitive the report in particular field is shown in this report using Impact Factor score.

However, Impact Factor cannot be a perfect indica-

tor to evaluate journals. Several drawbacks have been raised about Impact Factor (Seglen, 1997; Vanclay, 2012). Firstly, Impact Factor could be manipulated considering review papers because review papers are more quoted than any other articles (Simon, 2008). The biggest problem, on the other hand, is that it is hard to compare journals belonging to different subject categories by Impact Factor because there is average difference in score of Impact Factor in different subject categories. It is the result of deviations between different research areas as a consequence of various natures of their academic environments. For instance, social science researchers mostly prefer to publish books rather than journals but computer scientists prefer to present their results in conference proceedings. (Chen and Konstan, 2010). In addition, Impact Factor can be affected by the imbalance of the number of researchers and the characteristics of the journal publication system, such as the review process of each research field. Therefore, it cannot be a problem to compare researchers or research performance in the same field using Impact Factor, but it can cause problems when comparing different fields. In particular, Impact Factor of articles in the field of "Medical Science," and "Biology" are higher than any other field because the quotation is common. Thus, Impact Factor of low level journals in the field of "Medical Science," and "Biology" are sometimes higher than high level of those in the field of "Mathematics".

This drawback of Impact Factor is not a problem when evaluating researchers in the same field. However, when evaluating employees in universities or research institutes, it is common to evaluate researchers in other fields by applying the same criteria. In particular, Impact Factor based assessments can be discriminated against by differences between disciplines. Despite this problem, we mostly evaluate researchers by using Impact Factor in Korea. As mentioned above, Impact Factor is not a perfect evaluation because some researchers can get disadvantages in peculiar fields of study.

In previous studies, adjusted

Impact Factors for normalizing inter-field Impact Factor have been proposed (Pyo et al., 2016). There are many ways to normalize Impact Factor, but if journals belong to multiple categories, their relative rankings can be changed by the adjusted Impact Factor.

Our study focuses on introducing a new robust journal evaluation metrics that can normalize the differences in impact factor among various categories while maintaining the relative rankings of individual journals. In order to achieve this purpose, we use an optimization algorithm. As a result of applying the proposed method to the recent journal Impact Factor, it is confirmed that the error decreases.

The rest of this paper is organized as follows. In Section 2, we introduce well-known journal evaluation metric. In Section 3, we will introduce the methodology of

our research. In Sections 4 and 5, experimental result and conclusion are provided.

2. REVIEW OF WELL-KNOWN METRIC

To alleviate the problem that Impact Factors are different between subject categories, in previous studies, revised Impact Factor that divides Impact Factor with each representative was used (Pyo et al., 2016). Each representative values is obtained for each subject category. Each journal indexed in JCR is assigned to at least one subject category indicating a general area of science (SCI) or the social sciences (SSCI). There are about 230 subject categories in SCI and SSCI. Pyo et al. (2016) used Impact Factor weighted (number of articles) mean(AIF), a higher 20% ranked average Impact Factor, a higher 30% ranked average Impact Factor, a higher 50% ranked average Impact Factor, a higher 75% ranked average Impact Factor and total average Impact Factor as representative for each revised Impact Factor. An AIF, one of the representatives, indicates the average number of citations of articles in the subject category. Dividing Impact Factor as a representative value has the effect of normalizing the IF of each category.

These revised Impact Factors, however, have a problem when certain journal is related to more than two subject categories. In this case, we could not decide what we use as study representative. To deal with this problem, in the previous research they proposed two methods that using Average Impact Factor and Maximum Adjusted Impact Factor.

The suggested adjusted IF (A-IF) can reflect the information of each category and is defined by an average of impact factors divided by the AIFs of the included subject categories. Let $j^k \in J$ be the kth journal where $k \in \{1, \ldots\}$ in the alphabetical order and IF^k be the impact factor corresponding to journal j^k. Let $c_a \in C$ be the ath subject category where $a \in \{1, 2, \ldots, m\}$ in the alphabetical order and AIF_{c_a} is an aggregate impact factor corresponding subject categories. $j^k_{c_a, c_b}$ represents the kth journal included in subject categories c_a and c_b. C_k can be defined as a set of subject categories including the kth journal. A-IF of a journal, $j^k_{c_a, c_b, c_c}$ is the average value of the impact factor divided by each aggregate impact factor for a subject category included. $card(C_k)$ is the cardinality of C_k.

$$A - IF\left(j^k_{c_a, c_b, \ldots, c_l \in C_k}\right) = Average\left(\frac{IF^k}{AIF_{c_a}}, \frac{IF^k}{AIF_{c_b}}, \ldots, \frac{IF^k}{AIF_{c_l}}\right)$$

$$= \frac{1}{card(C_k)}\left(\sum_{n=c_a, c_b, \ldots, c_l \in C_k} \frac{IF^k}{AIF_n}\right)$$

Pyo et al. (2016) also suggested a way to use the

mean value of the quantiles of the impact factors of the journals listed in a subject category as a representative of journal's impact factor. The first new evaluation metric, namely QAVG-IF, is defined as follows

$$QAVG - IF_q \left(j^k_{c_a, c_b, \ldots, c_l \in C_k} \right)$$

$$= Average \left(\frac{IF^k}{AVG[Quan_q(c_a)]}, \frac{IF^k}{AVG[Quan_q(c_b)]}, \ldots, \frac{IF^k}{AVG[Quan_q(c_l)]} \right)$$

$$= \frac{1}{card(C_k)} \left(\sum_{n=c_a, c_b, \ldots, c_l \in C_k} \frac{IF^k}{AVG[Quan_q(c_n)]} \right)$$

Specifically, let $Quan_q(c_a)$ be a top q % quantile in the order of impact factors of the journals in category c_a and $AVG[Quan_q(c_a)]$ be the average impact factor for the journals included in top q % quantile in category c_a.

The second evaluation metric, namely QMAX-IF, uses a maximum value instead of using the average value.

$$QMAX - IF_q \left(j^k_{c_a, c_b, \ldots, c_l \in C_k} \right)$$

$$= MAX \left(\frac{IF^k}{AVG[Quan_q(c_a)]}, \frac{IF^k}{AVG[Quan_q(c_b)]}, \ldots, \frac{IF^k}{AVG[Quan_q(c_l)]} \right)$$

Although the above methods have the effect of normalizing Impact Factors of journals in more than one category, there is a problem that the existing ranking may be reversed. In previous research, they addressed this problem as Error Rate. Each method has six percent to ten percent Error rate.

To solve this problem we use F min search to find minimal Error Rate by using linearly combined 11 revised Impact Factor.

3. PROPOSED METHOD

In this research, we focused on finding a new measure that can replace original impact factor while maintaining ranking of individual journals as much as possible. So we tried to search the best linear combination of our features, which we made by manipulating original impact factor. So the new factor is defined as follows.

$$y(A) = A^T X,$$
$$\left\{ A = (a_1, a_2, \ldots, a_m)^T, \ X = (x_1, x_2, \ldots, x_m)^T \right\}$$

where A is a weight set, a_i indicates weights of each factors, and x_i indicates impact factors. In our situation, we have 11 revised Impact Factor so m = 11.

To estimate the weights of new Impact Factor, an objective function is required. Thus we defined a new variable, r(A), which we have to minimize.

Initially: r(A) = 0,

For f in every field:
  For all pair of two journals belonging to a field f:
    {If the relative rankings between journals have changed, i.e. original impact factor and y(A) show different results about which journal is better,
    r(A) = r(A)+1 }

Algorithm 1. Pseudo code for ranking check algorithm.

Our goal is to find appropriate weight set A which minimizes r(A). In this case, the objective function we should minimize is the number of times the ranking of any pair of two journals belonging to the same field has changed, so it is difficult to use analytical or numerical methods related to gradients. Thus we used F min search method with different initial values to search for the best coefficients of our new factor.

3.1 F Min Search

F min search is a simplex search method (It is also called the Nelder-Mead method) which means it is a direct search method, and it does not use numerical/ analytical gradients to compute optimized solution. So this algorithm is applied to nonlinear optimization problems, that derivatives may not be known. It is a heuristic search method that can converge to non-stationary points (McKinnon, 1999). In our case, the objective function is not differentiable, so it is appropriate to use F min search algorithm to solve this optimization problem.

To understand the algorithm, the concept of a simplex is fundamental. If there is an n-dimensional vector x, the simplex will be a special polytope made by n+1 vertices in n dimensions. For example, the simplex will be a triangle on a plane, and will be a tetrahedron in three-dimensional space. General idea of F min search is simple. In each search step, we pick new points around or inside the simplex. Then we evaluate the function values at those points and compare them to the function values at vertices. And if there's an improvement, we replace one of the vertices with one of our newly picked points. So new simplex is now generated. This iteration is repeated until the diameter of the simplex becomes smaller than specific tolerance we chose. In the next section, details of the algorithm will be described.

3.2 Algorithm of F Min Search

The following is typical F min search algorithm. We are trying to minimize the error rate r(A) where $A \in \mathbb{R}^n$, n = 11

1. A_i: List of points in the current simplex, $i = 1, \ldots, n+1$

2. Order the points in the simplex by function value, from the lowest $r(A_1)$ to the highest $r(A_{n+1})$. Get rid of A_{n+1} which has the worst function value (because we are finding minimum values) and add new point to the simplex. [Or we can replace n points except A_1, as we can see in Step 7.]

3. Generate the reflected point.

$$A_r = 2m - A_{n+1},$$

where

$$m = \Sigma n, i = \quad n1$$

and then calculate $r(A_r)$.

4. If $r(A_1) \le r(A_r) < r(A_n)$, accept A_r and iteration is terminated. **Reflect**

5. If $r(A_r) < r(A_1)$, calculate A_e which is expansion point.

$$A_e = m + 2(m - A_{n+1}),$$

and calculate $r(A_e)$.

   a. If $r(A_e) < r(A_r)$, accept A_e and iteration is terminated. **Expand**

   b. Otherwise, accept A_r and iteration is terminated. **Reflect**

6. If $r(A_r) \ge r(A_n)$ contraction is performed between m and the better of $\{A_{n+1}, A_r\}$.

   a. If $r(A_r) < r(A_{n+1})$ (i.e., A_r is better than A_{n+1}), calculate the below.

$$A_c = m + (A_r - m)/2$$

   and calculate $r(A_c)$. . If $r(A_c) < r(A_r)$, accept A_c and iteration is terminated. **Contract outside** Otherwise, continue with Step 7.

   b. If $r(A_r) \ge r(A_{n+1})$, calculate the below.
$$A_{cc} = m + (A_{n+1} - m)/2$$
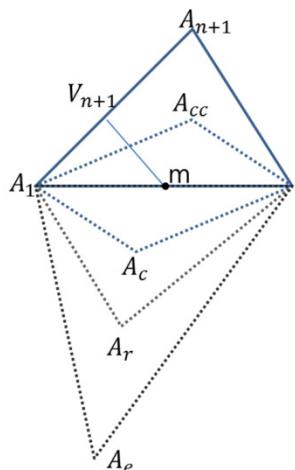
If $r(A_{cc}) < r(A_{n+1})$, accept A_{cc} and iteration is terminated. **Contract inside** Otherwise, continue with Step 7.

7. Calculate the n points.

$$V_i = A_1 + (A_i - A_1)/2$$

and calculate $r(V_i)(i = 2, ..., n+1)$. The simplex of next iteration is consisted of $A_1, V_2, ..., V_{n+1}$. **Shrink**

Figure 1 is an example of F min search procedure. Bold outline is the original simplex, and iteration continues until the simplex reaches the stopping criterion.

When applying the algorithm, the initial simplex is important. If the initial simplex is too small, there is a high risk of being stuck easily.

4. EXPERIMENTAL RESULTS

We have applied the proposed method to real data to evaluate the performance of the proposed method. For empirical analysis, we gathered 11,619 journal Impact Factor data of 2013, which are available from http://www.webofknowledge.com. Impact Factors are calculated yearly for academic journals that are listed in the Journal Citation Reports (JCR).

Table 1 shows number of subject categories and journals in 2013 JCR. JCR categorizes journals as SCI and SSCI according to the field. Some journals in the fields of psychiatry, psychology, and healthcare administration are included in both SCI and SSCI. There are also some journals belonging to more than two different categories depending on the subject of the journal.

Table 2 shows statistics well-known journal evaluation metrics proposed in Pyo et al. (2016). These statistics are about the mean values of each subject category. Most revised Impact Factors show less volatility than the original IF's. Small standard deviations mean that the variance between different subject categories has been reduced.

We can identify that optimal value depends on the initial value of the experiment. Since our optimization problem is not a convex optimization problem, it may have regarded local optimal value as global optimal value. We have experimented with various initial values. Table 3 shows the mean and standard deviation of the coefficients from the experimental results using randomly selected 11 different initial values.

Figure 1. Simple example of F min search.

Table 1. The number of subject categories and Journals in 2013 JCR

	# of subject categories	# of journals
SCI	176	8539
SSCI	56	3080

Table 2. Statistics of well-known evaluation metrics

	Mean	Std.	Skewness	Max-Min	CV
IF	1.816	0.8914	0.8148	4.620	49.0859
A-IF	0.818	0.1191	-0.1203	0.667	14.5599
$QAVG_{20}$ - IF	0.412	0.0546	-0.7413	0.396	13.2524
$QAVG_{30}$ – IF	0.484	0.0593	-0.8529	0.396	12.2521
$QAVG_{50}$ – IF	0.61	0.0698	-0.9067	0.482	11.4426
$QAVG_{75}$ – IF	0.762	0.0855	-0.8578	0.619	11.2205
$QAVG_{100}$ – IF	0.95	0.1051	-0.7781	0.789	11.0632
$QMAX_{20}$ - IF	0.505	0.0622	0.3752	0.410	12.3168
$QMAX_{30}$ - IF	0.584	0.0685	0.7001	0.423	11.7295
$QMAX_{50}$ - IF	0.729	0.0823	1.203	0.459	11.2894
$QMAX_{75}$ - IF	0.904	0.1025	1.467	0.580	11.3385
$QMAX_{100}$ - IF	1.127	0.1259	1.6642	0.660	11.1713

According to the results of Table 3, only a few of the 11 coefficients are statistically significant, but most of them are not significant. So we analyzed the results that showed the best performance rather than analyzing the effects of the coefficients. The best result is shown in Table 4.

Table 3. Statistics of coefficients

Mean	Std
0.6026812	1.1007351
-2.1344667	2.9427815
-0.7217748	0.9782207
-0.7389540	1.6901278
-0.1338603	0.4258694
1.2514219	1.5314143
-0.1772683	1.1677209
2.3245346	2.0691580
0.0992903	0.2737225
3.0543565	3.9370936
-0.1645822	0.8285788

Table 4. Best coefficient set

Initial value	Final value
0.0	0.1929779
1.0	-0.7076384
0.0	-0.4391840
1.0	0.0920815
0.0	0.4582388
0.0	-0.050070
0.5	-0.1717582
0.0	0.2123713
0.1	0.0098895
0.0	0.3508775
0.0	0.0200179

Our objective is to find the optimal adjusted Impact Factor that minimizes error rate while reducing the difference between categories. Thus we evaluate the proposed method in two respects; error rate and variance between subject categories.

We can see the error rate from various adjusted Impact Factors in Table 5. We found that our suggested adjusted Impact Factor has lower error rate compared to previous adjusted Impact Factors. The minimum error rate of the existing adjusted Impact Factors is 5.88% and the error rate of our suggested adjusted Impact Factor is 5.04%.

In Figure 2, we compared ordering error counts of 12 metrics. It is the number of combinations of journals whose relative rank reversed compared to original ranks. Our suggested Impact Factor results in 54,130 error counts and it reduces about 10,000 counts compared to the minimum error count of the existing adjusted Impact Factor. Considering the number of papers published in these journals, these results can be regarded as a remarkable reduction. For reference, the difference in the number of combinations of papers is about 1,600,000.

Table 5. Error rate from various adjusted Impact Factors

Adjusted IF	A-IF	QAVG-IF (20%)	QMAX-IF (20%)
Error Rate	6.24%	6.98%	8.17%
Adjusted IF	QMAX-IF (50%)	QMAX-IF (75%)	**QMAX-IF (75%)**
Error Rate	7.24%	5.94%	6.96%
Adjusted IF	QAVG-IF (30%)	QMAX-IF (30%)	**QAVG-IF (50%)**
Error Rate	6.56%	7.69%	6.14%
Adjusted IF	QAVG-IF (100%)	QMAX-IF (100%)	Adj-IF
Error Rate	5.88%	6.92%	**5.04%**
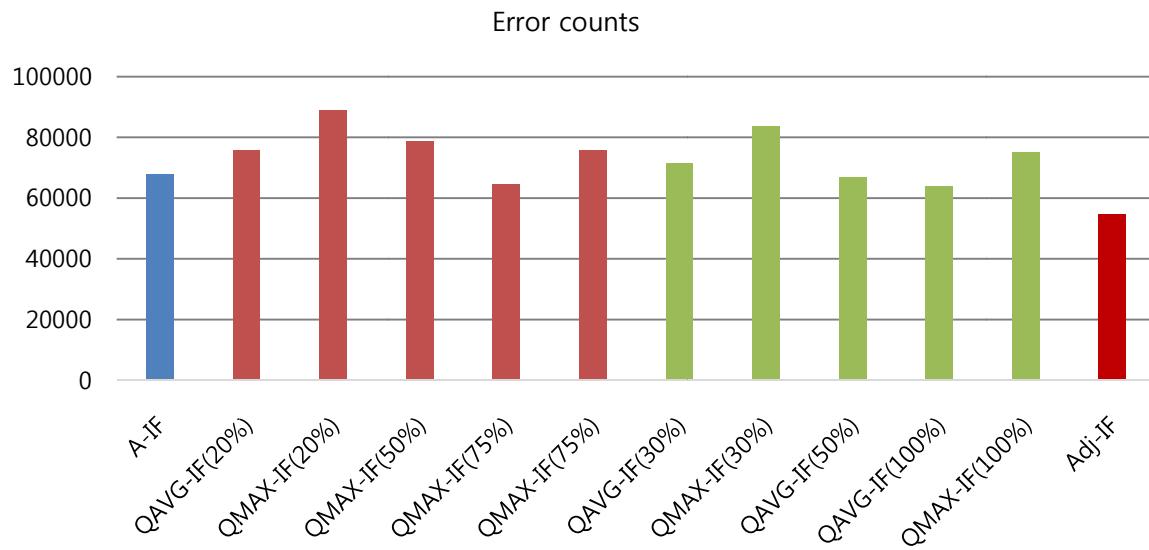
Figure 2. Ordering error counts of 12 metrics.
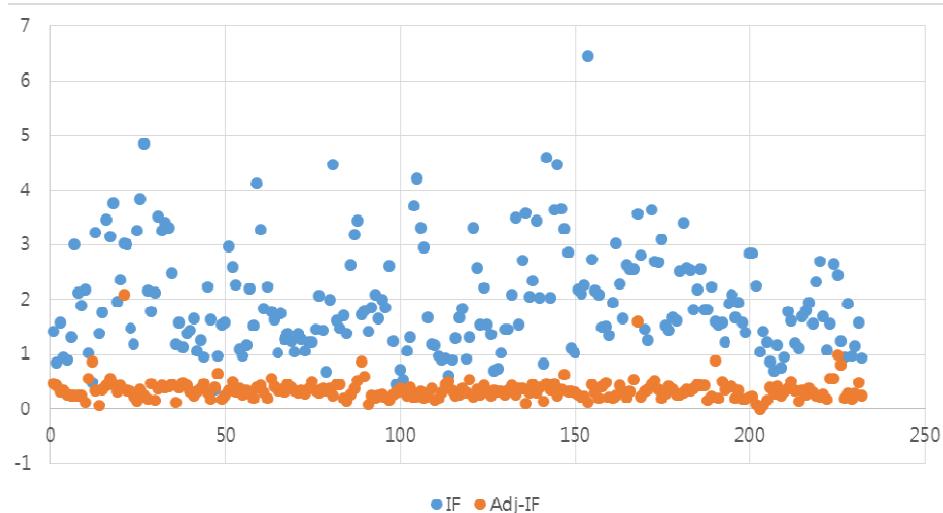
Figure 3. Average IF by categories.

Secondly, what we need to look at is how much new adjusted Impact Factors reduce cross-sector volatility. Table 6 shows statistics of proposed adjusted Impact Factor. As shown in Table 6, the standard deviation value of the proposed Impact Factor is the smallest when compared with the existing values. The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, shows that the proposed method has the smallest variation (10.5134%) in comparison with other methods.

In Figure 3, we can compare average Impact Factor with average adjusted Impact Factor by categories' distribution. From Figure 3 we can identify that our suggested adjusted Impact factor has reduced Impact factor variance between different categories.

Figure 4 compared average Impact factor and average adjusted Impact Factor in categories related 'Biology' and 'Engineering'. We found that our adjusted Impact Factor is more resonable journal evaluation indicator in aspects of variance of categories.

Table 6. Statistics of proposed Adj_IF

| Mean	Std.	Skewness	Max-Min	CV
0.409	0.0430	1.102	0.246	10.5134
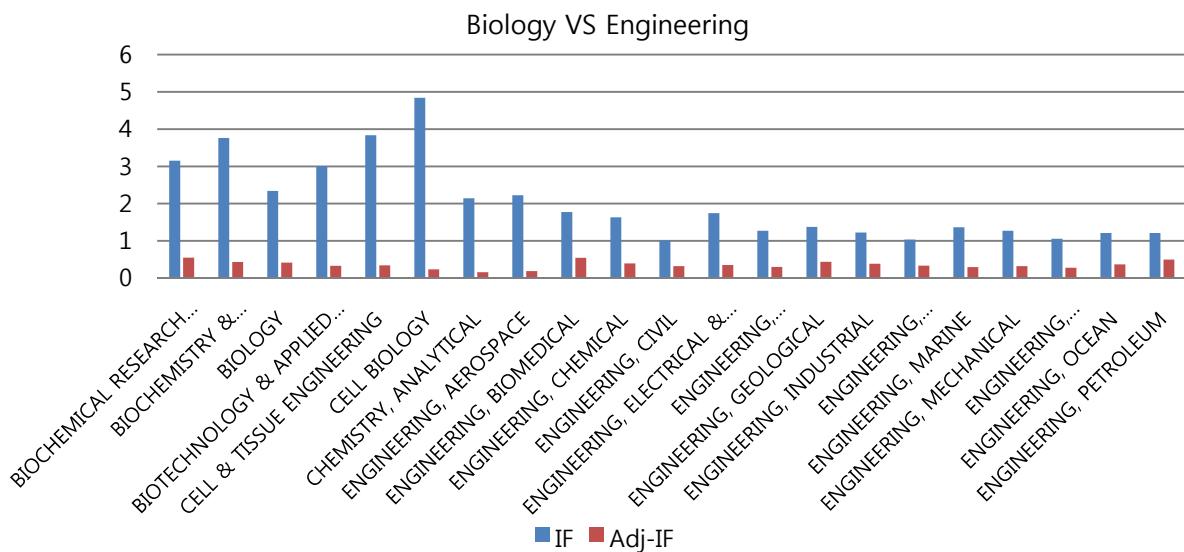
Figure 4. Average impact factor in biology and engineering.

5. CONCLUSIONS

It is important to evaluate research performance of researchers. Impact Factor has been widely used as an evaluation index. In this study, we propose a method to minimize the error that Impact Factor ranking is changed by finding the optimal combination of the modified Impact Factors to solve the imbalance problem between the research fields of Impact Factor. As a result, the deviation by subject category has been reduced and the error rate has been reduced compared with the existing method. Therefore, the number of damaged journals and papers affected by the change in rankings has been significantly reduced By using F min search method to find optimal value, our model showed error rate of 5.04%, while previous model showed 6.79% on average. Since F min search algorithm is sensitive to the initial value, there is room to improve the results by studying how to set good initial values in further research. We need to apply various algorithms and object functions to find the optimal journal evaluation measure in future study.

Although there are limitations left in evaluating research performance using Impact Factor, the proposed adjusted Impact Factor is expected to be helpful in evaluating researchers from different disciplines.

ACKNOWLEDGEMENTS

REFERENCES

Chen, J. and Konstan, J. A. (2010), Conference paper selectivity and impact, *Communications of the ACM*, **53**(6), 79-83.

McKinnon, K. I. M. (1999), Convergence of the Nelder–Mead simplex method to a nonstationary point, *SIAM Journal of Optimization*, **9**(1), 148-158.

Pyo, S., Lee, W., and Lee, J., (2016), A Novel journal evaluation metric that adjusts the impact factors across different subject categories, *Industrial Engineering Management Systems*, **15**(1), 99-109

Sahel, J. A. (2011), Quality versus quantity: Assessing individual research performance, *Science Translational Medicine*, **3**(84), 84cm13.

Seglen, P. O. (1997), Why the impact factor of journals should not be used for evaluating research, *BMJ*, **314**(7079), 498-502.

Simons, K. (2008), The misused impact factor, *Science*, **322**(5899), 165.

Vanclay, J. K. (2012), Impact factor: Outdated artifact or stepping-stone to journal certification?, *Scientometrics*, **92**(2), 211-238.