2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

RATE: a Review of Reviewers in a Manuscript Review Process

Wei Wei, Kam Tong Chan, Irwin King, Jimmy Ho-Man Lee
Department of Computer Science & Engineering
The Chinese University of Hong Kong
{wwei, ktchan, king, jlee}@cse.cuhk.edu.hk

Abstract

In this paper, we propose a novel approach called, Reviewers Authority Testing and Evaluation (RATE), to improve the effectiveness of a manuscript review process. In the proposed RATE approach, we define a RATE model to express a manuscript review process mathematically. We then design a RATE algorithm to rank the authority of each reviewer in the RATE model and consequently calculate the quality score for each manuscript. The experimental results demonstrate that the performance of the RATE algorithm is superior to existing approaches. Furthermore, the experiments on testing algorithm's parameter settings also demonstrate that the proposed RATE algorithm behaves effectively and stably.

1 Introduction

As a rule, every manuscript review process is presented two essential problems. One problem is how to assign submitted manuscripts to the most appropriate reviewers. Although several methods are proposed on this issue [2, 3, 1], the manuscript-reviewer assignment procedure is usually not perfect in practice. The other problem is how to summarize all the ratings of reviewers to best uncover the truthful score of each submitted manuscript. Our work focuses on solving the second problem involved in a good review process by authority ranking for each reviewers.

The concept of authority ranking has been widely recognized in the field of information retrieval. HITS algorithm [4] has been proposed to distinguish *authorities* and *hubs* in the web-searching process. An E-learning platform [7] was reported to evaluate the users' authority to improve the learning efficiency in the online discussion process. Two algorithms on ranking reviewers' authority in a manuscript review process have been proposed by Riggs [6] and Lauw [5]. Although the experiments on both existing algorithms indicate positive impact of ranking reviewers' authority in a manuscripts review process, we could still find improvements to these techniques, since both of the algorithms do not show enough recognition on the difference between consistent inaccuracy from inconsistent inaccuracy. Intuitively, we believe that a reviewer with more inconsistent inaccuracy is worse than another reviewer with more consistent inaccuracy but less inconsistent inaccuracy. When the two types of inaccuracies are accounted, they should not only be counted differently in quantity but also be counted by different variables.

In this paper, we propose a novel approach, called Reviewers Authority Testing and Estimation (RATE) to solve the second problem involved in a good review process. In Section 2, we formulate the proposed RATE approach in details. In Section 3, we present the experimental results. Finally, in Section 4, we conclude this paper.

2 The RATE Approach

Definition 1 (RATE Model) *A* ***RATE Model*** *is a octad group $T = (R, M, C, E, W, S, \Omega, \Sigma)$. R is a set of reviewers in a review process. M is a set of manuscripts to be reviewed. C is a set of optional ratings from which a reviewer may choose to rate a manuscript. E is a relation set which describe the ratings the manuscripts obtain and the reviewed relations between manuscripts and their reviewers: $E = \{e_{rm} | r \in R, m \in M, e_{rm} \in C$ and a manuscript m is rated by a reviewer r with $e_{rm}\}$. W is a set of values which are respectively corresponding to every reviewer's authority weight in the manuscript review process. S is a set of result scores of manuscripts: $S = \{s_m | s_m$ is calculated by function $\Sigma\}$. Ω is a mapping function which maps a rating to a corresponding numerical value. Σ is a function of summarizing a manuscript's reviewers' ratings.*

2.1 The Ω Function Design

The Ω function in the RATE model basically serves as a mapping function from reviewers' literal ratings to appropriate numerical values in the range of $[0, 1]$. In order to make the ratings into two attitude groups, each of which respectively favors one of extreme outcomes of a review process, a critical value θ and a cushion radius σ are specified in the Ω function, which will divide the range of $[0, 1]$ into three scales,i.e. $[0, \theta - \sigma]$, $[\theta - \sigma, \theta + \sigma]$, and $[\theta + \sigma, 1]$.

In the RATE model, we suggest to set the critical value $\theta = 0.5$ and the cushion radius $\sigma = 0.1$. An example of an Ω function of mapping the above five ratings to corresponding numerical values is shown in Table 1.

978-0-7695-3496-1/08 $25.00 © 2008 IEEE
DOI 10.1109/WIIAT.2008.42

Table 1: An Example of The Ω Function

Ratings	Numerical Values	Attitude Groups
Strong Accept	1	Favor to Accept
Weak Accept	0.7	Favor to Accept
Neutral	0.5	$\theta = 0.5$ and $\sigma = 0.1$
Weak Reject	0.3	Favor to Reject
Strong Reject	0	Favor to Reject

2.2 The Σ Function Design

The Σ function shown as Eq. 1 in the RATE model serves to summarize all the reviewers' ratings for a manuscript into an overall rating score.

$$s_m = \frac{\sum\limits_{r \in R_m} w_r \cdot \Omega(e_{rm})}{\sum\limits_{r \in R_m} w_r}, \tag{1}$$

where s_m is an overall rating score summarized from all m's reviewers, R_m is a set of reviewers for the manuscript m, w_r is the authority weight of the reviewer r, e_{rm} is a rating the reviewer r give to the manuscript m and $\Omega(.)$ is the rating-value mapping function used in the RATE model.

2.3 Calculation of W

2.3.1 Rating inaccuracy

The method to compute the **rating inaccuracy** of each reviewer is different from existing approaches [6, 5]. Both of the existing approaches failed to differentiate two types of reviewers' inaccuracies which are respectively named **consistent inaccuracy** and **inconsistent inaccuracy** in the proposed RATE approach.

Consistent inaccuracy The consistent inaccuracy indicates a reviewer's inaccuracy in rating degree but not in rating attitude. The formula of calculating the consistent inaccuracy of a reviewer r is as follow:

$$x^{<r>} = \frac{\sum\limits_{\forall m \in M'_r} |s_m - \Omega(e_{rm})|}{n'_r}, \tag{2}$$

where $x^{<r>}$ is a reviewer r's consistent inaccuracy, M'_r is a set of manuscripts whose ultimate average ratings are considered to be consistent with the reviewer r's ratings in attitude, m represents one manuscript in set M'_r, e_{rm} is the rating the reviewer r give to the manuscript m, Ω is the mapping function of the model RATE and n'_r is the cardinality of the set M'_r.

Inconsistent inaccuracy The inconsistent inaccuracy indicates a reviewer's inaccuracy in the rating attitude. The formula of computing the inconsistent inaccuracy of a reviewer r is as follows:

$$y^{<r>} = \sum_{\forall m \in M''_r} \frac{|s_m - \Omega(e_{rm})|}{n''_r}, \tag{3}$$

where $y^{<r>}$ is a reviewer r's inconsistent inaccuracy, $M_r^{''}$ is a set of manuscripts whose ultimate average ratings are considered to be inconsistent with the reviewer r's ratings in attitude, m represents one manuscript in set $M_r^{''}$, e_{rm} is the rating the reviewer r give to the manuscript m, Ω is the mapping function of the RATE model and n''_r is the cardinality of the set M''_r.

Algorithm 1 Consistent and inconsistent inaccuracy calculation

INPUT:
M_r: a set of manuscripts reviewed by r
$\{e_{rm}\}$: a set of ratings r give to manuscripts in M_r
$\{s_m\}$: a set of average rating scores of manuscripts in M_r
Ω: a Ω function of the RATE model
θ: a specified critical value of Ω function
σ: a specified cushion radius of Ω function
OUTPUT:
$x^{<r>}$: the reviewer r's consistent inaccuracy
$y^{<r>}$: the reviewer r's inconsistent inaccuracy
BEGIN
for all m in M_r **do**
  if ($s_m \in [0, \theta - \sigma]$ and $\Omega(e_{rm}) \in [\theta + \sigma, 1]$) or
   ($s_m \in [\theta + \sigma, 1]$ and $\Omega(e_{rm}) \in [0, \theta - \sigma]$) **then**
   e_{rm} induces inconsistent inaccuracy recorded by $y^{<r>}$
  else
   e_{rm} induces consistent inaccuracy recorded by $x^{<r>}$
  end if
end for
RETURN $(x^{<r>}, y^{<r>})$
END

Calculation of rating inaccuracy Rating inaccuracy is a linear combination of a reviewer's consistent inaccuracy and inconsistent inaccuracy. We first utilize Algorithm 1 to differentiate consistent inaccuracy and inconsistent inaccuracy. Then we merge a reviewer's $x^{<r>}$ and $y^{<r>}$ into the reviewer's rating inaccuracy by a linear combination with formula as follows:

$$z^{<r>} = \beta \cdot x^{<r>} + (1 - \beta) \cdot y^{<r>} (0 \leqslant \beta \leqslant 1). \tag{4}$$

In Eq. (4), β is a proportion adjustment between consistent inaccuracy and inconsistent inaccuracy. When applied in RATE algorithm, the value of $z^{<r>}$ of each reviewer r will be normalized into a range of $[0, 1]$.Reviewers' authority weights are calculated by their rating inaccuracies with the intuition that reviewers' authority weights are in inverse ratio of their rating inaccuracies. Therefore we use the following formula to compute a reviewer's authority weight.

$$w^{<r>} = 1 - z^{<r>}. \tag{5}$$

Specifically, we can notice that if we set $\theta = 0$, $\sigma = 0$, and $\beta = 1$, the process of calculating the authority

Algorithm 2 RATE Algorithm

INPUT:
R: a set of reviewers
M: a set of manuscripts
E: a relation set describe review relations and ratings
Ω: a Ω function, e.g., Table 1
Σ: a Σ function, e.g., Eq. (1)
θ: a specified critical value of Ω function
σ: a specified cushion radius of Ω function
t: a specified number of iterations
OUTPUT:
W: a set of reviewers' authority weights
S: a set of manuscripts' average rating scores
PARAMETER:
β: a parameter for linear combination in Eq. (4)
VARIABLE:
x: a vector of $x^{<r>}$ for every reviewers
y: a vector of $y^{<r>}$ for every reviewers
z: a vector of $z^{<r>}$ for every reviewers
BEGIN
Initialize elements in W to be 1
for all m in M **do**
    Compute s_m with Σ function
end for
while Algorithm Not Converge and less than t iterations **do**
    for all r in R **do**
        Compute $x^{<r>}$ and $y^{<r>}$ with Algorithm 1
    end for
    for all r in R **do**
        Compute $z^{<r>}$ with Eq. (4)
    end for
    for all r in R **do**
        Compute $w^{<r>}$ with Eq. (5)
    end for
    Normalize elements w in set W
    for all m in M **do**
        Compute s_m with Σ function
    end for
end while
RETURN (S, W)
END

weights for every reviewers is exactly as same as Riggs's method [6], which indicates the generalization of our proposed approach.

2.4 Rate Algorithm – Calculation of S

The RATE algorithm is an iterative process of computing the average rating score of every manuscript. In each iteration, the algorithm utilizes the reviewers' authority weights obtained in the last iteration to calculate the current average rating score of every manuscript and then evaluates every reviewer's authority weight in the current iteration. The algorithm terminates if every reviewer's authority weight remains unchanged or t rounds of iterations is finished. The details of the RATE algorithm is presented in Algorithm 2.

3 Experiments

The experiments on algorithms of authority ranking are always challenging tasks. In order to better evaluate the algorithm's performance, we design a program to simulate the manuscript review process so that we can prefabricate ground truth of the generated data for our experiments.

3.1 Simulation data sets and evaluation metrics

The manuscripts in the simulation program are classified into five subsets ($M^{[5]}$, $M^{[4]}$, $M^{[3]}$, $M^{[2]}$, $M^{[1]}$) according to the predefined ground truth of their qualities("Strong Accept", "Weak Accept", "Neutral", "Weak Reject", "Strong Reject"). In order to measure the performance of the algorithm, we define the concept of **penalty value** of an algorithm with the following formula:

$$pen(al) = \sum_{s_m<\theta} \Big(\sum_{m\in M^{[5]}} (\theta - s_m)\cdot 1 + \sum_{m\in M^{[4]}} (\theta - s_m)\cdot 0.8\Big) + \sum_{s_m>\theta} \Big(\sum_{m\in M^{[1]}} (s_m - \theta)\cdot 1 + \sum_{m\in M^{[2]}} (s_m - \theta)\cdot 0.8\Big), \tag{6}$$

,where θ is the average score of all the evaluated manuscripts in the review process, and s_m is the rating score the evaluated algorithm assigns to a manuscript m. The penalty value of the process without considering reviewers authority ranking is denoted as $pen(os)$. The performance of the evaluated algorithm al is defined as:

$$PF(al) = \frac{pen(os) - pen(al)}{pen(os)}. \tag{7}$$

In the simulation program, we predefine 1000 manuscripts in M, i.e. each subset $M^{[\cdot]}$ contains 200 manuscripts, and 100 reviewers in R, i.e. 50 authoritative reviewers, 30 average reviewers, and 20 unreliable reviewers. In each round of simulation we let each manuscript in M be reviewed by three reviewers in R. We run the simulation program three times to generate 3 data sets on which we will test the proposed approach.

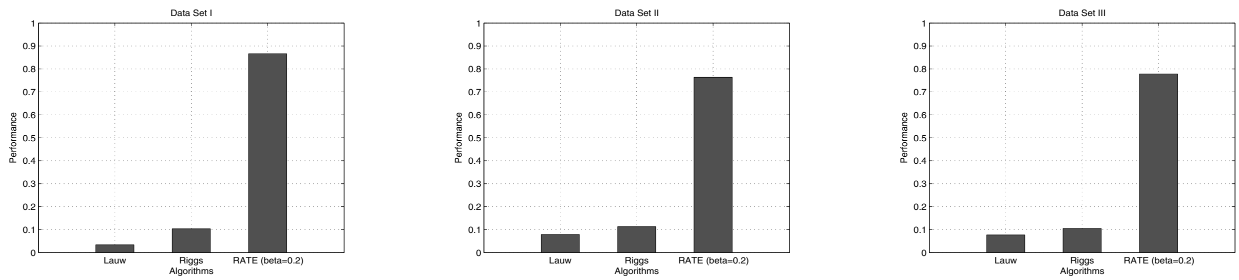
3.2 Experimental results

3.2.1 Performance comparisons

We conduct experiments on algorithms' performances evaluation and comparison on the three simulated data sets. The parameter β of the RATE algorithm in the comparison is set to be 0.2. The performance values for the three algorithms are respectively calculated and shown in the Fig. 1, which indicates that by using the RATE algorithm to rank each reviewer's authority the penalty value of the process greatly bottoms out.

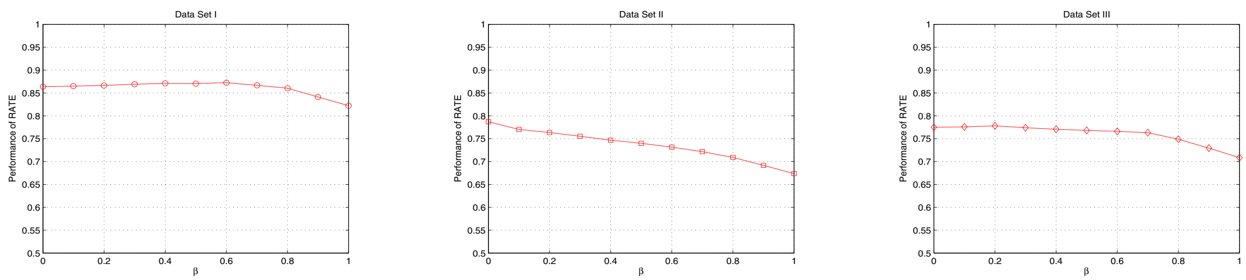
3.2.2 Impact of β

The parameter β plays an important role in the RATE approach. It serves as the adjustment factor between consistent inaccuracy and inconsistent inaccuracy when we calculate a reviewer's rating inaccuracy. We conduct experiments on examining the impact of β of the RATE algorithm



(a) (b) (c)

Figure 1: Algorithms Performances



(a) (b) (c)

Figure 2: Impact of β

and present the experimental results in Fig. 2. From the results in Fig. 2, we can see that the parameter β may affect the performance of the RATE algorithm to a certain extent. However, as the variation of the parameter the performances of the RATE algorithms are stable and better than the other two algorithms.

4 Conclusions

The effectiveness of a manuscript review process can be improved if each reviewer's authority can be taken into account when the TPC summarizes all the reviewers' ratings. The proposed RATE approach contains two parts in contributing to improvement of a manuscript review process. Firstly, a RATE model is introduced to express the manuscript review process mathematically. Secondly, a RATE algorithm is involved to rank the authority of reviewers and consequently best uncover the ground truth of the manuscripts' qualities. The experimental results compared with other two existing algorithms show that the performance of the RATE algorithm is superior. The experiments on the parameter β's testing indicate that the value of β may affect the RATE algorithm's performance in some degree and the RATE algorithm behaves stably as the variation of its parameter.

References

[1] C. Basu, H. Hirsh, W. W. Cohen, and C. G. Nevill-Manning. Technical paper recommendation: A study in combining multiple information sources. *J. Artif. Intell. Res. (JAIR)*, 14:231–252, 2001.

[2] S. T. Dumais and J. Nielsen. Automating the assignment of submitted manuscripts to reviewers. In *Research and Development in Information Retrieval*, pages 233–244, 1992.

[3] S. Hettich and M. J. Pazzani. Mining for proposal reviewers: lessons learned at the national science foundation. In *KDD*. ACM, 2006.

[4] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[5] H. W. Lauw, E.-P. Lim, and K. Wang. Summarizing review scores of unequal reviewers. In *SIAM International Conference on Data Mining*, 2007.

[6] Riggs, Tracy and Wilensky, Robert. An algorithm for automated rating of reviewers. In *JCDL'01: Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries*, Scholarly Communication and Digital Libraries.

[7] W. Wei, J. Lee, and I. King. Measuring credibility of users in an e-learning environment. In *Poster Proceedings of 16th International World Wide Web Conference, WWW 2007, Banff, Alberta, Canada*. ACM, 2007.