Collaborative Aggregation of Individual Rating for Group Evaluation

Li Yu  
Renmin University of China, buaayuli@ruc.edu.cn

Haiyan Zhang  
Renmin University of China, han.zhang@mgt.gatech.edu

Liang Xun  
Renmin University of China, xliang@ruc.edu.cn

Follow this and additional works at: http://aisel.aisnet.org/pacis2013

Recommended Citation
http://aisel.aisnet.org/pacis2013/63

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2013 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
COLLABORATIVE AGGREGATION OF INDIVIDUAL RATING FOR GROUP EVALUATION

Li Yu, School of Information, Renmin University of China, Beijing, P.R.China, State Key Laboratory of Digital Publishing Technology, Beijing, P.R.China, buaayuli@ruc.edu.cn

Haiyan Zhang, School of Information, Renmin University of China, Beijing, P.R.China, han.zhang@mgt.gatech.edu

Liang Xun, School of Information, Renmin University of China, Beijing, P.R.China, xliang@ruc.edu.cn

Abstract

The aggregation of individual evaluations into a group evaluation is a key issue in decision theory. Inspired by the collaboration idea on Web2.0, two group evaluation methods of collaborative weighting evaluators where evaluators rate the objects evaluated are presented to form group decision. Among the most commonly applied methods for group evaluation are to average the scores obtained by each objects evaluated and then consider the associated collective evaluation, which not consider the difference between individual evaluators. The proposed methods can offset the effect resulted by some evaluator’s irregular evaluation through considering the contribution of the individual evaluation to the collective evaluation. Sometimes the irregularity is owing to the subjectivity reason. For example, since some evaluators are prejudiced against some objects evaluated, the evaluators will subjectivity give very high or very low rating on the objects evaluated. In this paper, two nonlinear group evaluation methods achieved through mutual restraints between individual evaluators and consequently are more stable than traditional group evaluation method in an actual example and a synthetic data set.

Keywords: Web2.0, Group decision, Collaborative evaluation, Group agreement
1 INTRODUCTION

Evaluation is an important research field in decision science since evaluation result will have a direct impact on decision making. Also, evaluation is an important and difficult issue in management science, which always attracts a large number of scholars’ attention (Chen et al. 2004). Furthermore, evaluation can be widely applied many application areas, such as sports game ranking, qualification of college student grade point average, review of fund committee project, and recommendation based on evaluation of user to goods in e-commerce and etc. Different application areas have different significance.

Currently, most of the widely used evaluation methods are based on indicator system (Zhang et al. 2007), called as Evaluation based-Indicator (referred to as IE) method. Generally, the IE method needs to solve two key problems. The first one is to establish a reasonable evaluation indicator system where each indicator is relatively independent and can be measured. The second one is to determine the value of indicator for each evaluated object. IE methods are usually more adapted to evaluate structured objects that are measured objectively through a number of indicators (Lan et al. 2009). And specifically the following three conditions are often to be satisfied simultaneously: Firstly, a completely accurate evaluation index system is needed. Secondly, each indicator has a reasonable weight value. Thirdly, each indicator of evaluated object can be measured objectively. But in traditional evaluation method based on indicator system, the confirmation of indicator system is very difficult, and the indicator and weight value are easily influenced by experts. Especially in some major evaluation, an authoritative expert may play a leading role in the evaluation process and affect the evaluation of other experts. So IE methods is often difficult to be satisfied three important conditions simultaneously and thus IE methods are hard to be used under many situations. In this paper we mainly concern on evaluation without indicator but considering the difference between the individual and the collective evaluations.

In fact, the aggregation of individual evaluation into a group evaluation is also a typical group decision problem and has a wide range of applications in social choice and voting systems (I. C. 2011). Among the most commonly applied methods for group evaluation are to average the scores obtained by each alternative and then consider the associated collective evaluation (Dorit 2006), which not consider the difference between individual and individual evaluators. However there is some evaluator’s irregularities evaluation widely existing in management evaluation where evaluators rating the alternatives. Sometimes the irregularity is owing to the subjectivity reason. For example, since there is a good or bad friendship between evaluator and evaluated person, the evaluator will subjectivity give very high or very low evaluation on evaluated person. Sometimes the irregularities evaluation is resulted by the limited profession level of evaluators who are hard to objectively give correct evaluation. This class of procedures may include weighting factors given by experts in advance in order to emphasize the relative importance of some individuals. However, in this paper, we present two group evaluation methods to collaboratively find the evaluator’s irregularities and assign different weight for all evaluators to offset the effect resulted by some evaluator’s irregularities evaluation. Therefore two nonlinear group evaluation methods achieved through mutual restraints between individual evaluators and consequently are more stable than traditional group evaluation methods.

With the emergent and development of Web2.0 in recent years, it emerged as a new widely used internet mode, such as Blog, Tag, SNS, RSS and Wiki, etc. Its key idea is group collaboration, which have attracted more and more attention, and have been successfully applied in many fields such as social software, personalized recommendation (Breese et al. 1998). In fact, the ideas on group collaboration also play an important enlightening role in management science and group decision. A preliminary exploration on the application of collaboration in management science has been taken, and a new management theory - Collaborative Management theory, has been proposed (Yu et al. 2006). In essence, it is a specific application of Web2.0 idea on group collaboration in management science. Currently, the majority of evaluation criteria are very subjective. Sometimes, a certain evaluator, such as authoritative experts, may have critical impact on final evaluating result. According to the
collaboration idea of Web2.0, if the evaluators can be restricted with each other, then the issue can be avoided. Similar to this study, a kind of subject-object collaborative group evaluation method has been proposed (Zhang et al. 2010) which emphasizes the participation of the one being evaluated, and aggregate the evaluation results of both subject and object. This paper discusses how to apply group collaboration idea into group evaluation.

The paper is organized as follows. First, the collaborative evaluation theory is proposed in section 2. In section 3, two special collaborative evaluation methods for group evaluation are described in detail. In section 4, a series of experiments are conducted to verify the proposed methods. Finally, the conclusions are given in section 5.

2 COLLABORATIVE EVALUATION

This section outlines the novel idea of collaborative evaluation for group evaluation in decision theory.

2.1 Definition

Broadly speaking, if an evaluation method embodies some ideas on collaboration, then the method is a Collaborative Evaluation (short for CE) method. The word ‘Collaborative’ reflects mutual restriction among evaluators, and the evaluation results depend on all evaluators rather than just individual. CE is a kind of evaluation theory, rather than only a specific evaluation method. CE can have different implementation patterns according to different types of evaluation. Just like that Web2.0 is a Web mode, different applications of Web2.0 systems may vary. But no matter what kind of Web2.0 systems, they have common characteristics: Grass-roots, Decentralization and Collaboration, etc. Similarly, the most essential idea of CE is that evaluators constrained and influenced each other, and minimize individual evaluation subjective factor.

2.2 Characteristics

In general, CE has the following key characteristics:

First, the evaluation result is stable. The evaluation results would not easily be affected by one or a small number of evaluators. For example, in some situation, the evaluator is not very familiar with the evaluation areas or not very professional. Sometime, cheating deliberately in evaluation could happen because of the personal relationship between the evaluator and the evaluated person. Under these situations, CE still has the stable evaluation result since the personal evaluation does not make direct effect on the final result.

Secondly, an evaluation indicator system confirmed by experts in advance is not required. CE only needs the overall evaluation points while the evaluation indicator systems are not needed. Of course, evaluation indicator system is benefit for the overall evaluation.

Thirdly, the more evaluators are, the more accurate the evaluation results are. CE generally requires a number of evaluators. The more evaluators are, the higher the collaboration among the evaluators is. Then the evaluation systems are more stable, and the evaluation result is more accurate. In fact, suppose that n is the number of evaluators, then the number of communication channel between evaluators (collaboration among the evaluators) is n (n-1)/2. Obviously, if n increase, the evaluate accuracy will improve with quadratic level.

2.3 Condition for Application

According to above analysis, CE is mainly suitable for the following evaluation situations. First, the premise of CE is that each evaluator is able to make initial and comprehensive evaluation respectively. So CE applies to the fields that are easier to make overall evaluation. But inaccurate initial evaluation is allowed in CE. Secondly, it is easy for public to participate. Since CE requires a lot of evaluators, it
is critical that common users could be easily involved, such as through the Internet. Thirdly, it spends long time to finish evaluation task. Because CE needs more users, the entire evaluation duration is not easy to control, and a long time is needed.

2.4 Advantages and Disadvantages

CE has following advantages. First, its evaluation results are more accurate and more objective. In traditional evaluation method based on indicator system, the confirmation of indicator system is very difficult, and the indicator and weight value are easily influenced by experts. Especially in some major evaluation, an authoritative expert may play a leading role in the evaluation process and affect the evaluation of other experts. CE doesn't need an evaluation indicator system confirmed by experts in advance. Furthermore the more evaluators, the more accurate and objective the results are. Secondly, it could be applied to the evaluation of unstructured object. Because no indicator is needed, CE can be used to evaluate complex (unstructured) object which is difficult to be measured through some multi-attributes or indicator. Thirdly, there is a low requirement on evaluators. In traditional evaluation, the evaluators generally are a few experts. In collaborative evaluation, no special requirement is needed for evaluators. They could be expert, or common user.

CE also has following disadvantages. First, it cannot avoid group error in evaluation. CE method cannot deal with the same mistake made by all evaluators. Secondly, comprehensive evaluation for the object being evaluated is needed initially. But sometime it is difficult. Thirdly, in general, more evaluators and longer evaluation time are needed for CE, which may result in cost increase of evaluation.

3 COLLABORATIVE EVALUATION METHODS FOR AGGREGATING GROUP EVALUATION

In this section, we first formally describe the research question based on rating score of evaluators. After that, in section 3.2, we propose a collaborative evaluation method based on rating difference, called CE_DIFF. Then, in section 3.3, we detail state the other collaborative evaluation method based on overall agreement, called CE_AGREE.

3.1 Question Description

First, let’s give the evaluation description. Supposed that there are m objects evaluated by n evaluators \( U = \{ U_1, U_2, ..., U_n \} \), as shown in Table 1, where \( R_{ij} \) is evaluating score of \( U_i \) on the object \( O_j \). The question is how to sort the objects according to the evaluation on the object.

<table>
<thead>
<tr>
<th>( U_1 )</th>
<th>( O_1 )</th>
<th>( O_2 )</th>
<th>...</th>
<th>( O_m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U_2 )</td>
<td>( R_{11} )</td>
<td>( R_{12} )</td>
<td>...</td>
<td>( R_{1m} )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( U_n )</td>
<td>( R_{n1} )</td>
<td>( R_{n2} )</td>
<td>...</td>
<td>( R_{nm} )</td>
</tr>
</tbody>
</table>

Table 1. Evaluation Matrix

The simplest method is to average score of all evaluator’s evaluation as the final score according to the traditional evaluation method, the ranking of objects being evaluated are based on average score. But there are two shortcomings for averaging method. The first one is the practice favoritism and fraudulence. An evaluator did not play fair subjectively (cheating in evaluation) or is not professional (non-professional in non-professional) hardly to make accurate evaluation, and the final evaluation results would be affected directly. Another one is that weights of evaluators do not be considered.

3.2 Collaborative Evaluation Based on Rating Difference (CE_DIFF)
In this part, Collaborative Evaluation Based on Rating Difference (CE_DIFF) is proposed. Different from traditional IE method, CE_DIFF assigns different weight to different evaluator. In detail, larger weight was assigned to the evaluator who makes more accurate evaluation, and accuracy of the evaluation for evaluator was computed according the evaluations errors between the evaluator and others. Specially, the evaluation is inaccurate if the evaluation error is big, and correspondingly less weight is assigned to the evaluator. That is to say, whether the evaluation is accurate or not depend on all evaluators not be decided by someone. In the method, the ideas on collaboration and decentralization are fully embodied. In practice, CE can be computed as following steps.

Step 1: For any evaluated object \( O_j \), computing average score for each evaluated object,

\[
\frac{1}{n} \sum_{i=1}^{n} R_{ij} \quad (j=1, 2, ..., m)
\]

Step 2: For any evaluated object \( O_j \), computing the difference between average score and evaluating score given by evaluator \( U_i \),

\[
V_i = |R_i - R_j| = |R_i - \frac{1}{n} \sum_{j=1}^{m} R_{ij}| \quad (i=1,2,\ldots,n; \ j=1,2,\ldots,m)
\]

Step 3: Computing the weights of the evaluator based on evaluating difference,

\[
W_i' = S - V_i = S - |R_i - \frac{1}{n} \sum_{j=1}^{m} R_{ij}|
\]

where \( S \) means evaluation scale, in the following example, \( S=100 \).

Step 4: Normalize the evaluating weight.

\[
W_i = \frac{W_i'}{\sum_{i=1}^{n} W_i'} = \frac{S - |R_i - \frac{1}{n} \sum_{j=1}^{m} R_{ij}|}{\sum_{i=1}^{n} (S - |R_i - \frac{1}{n} \sum_{j=1}^{m} R_{ij}|)}
\]

Step 5: Computing final evaluation score,

\[
R_{ij} = \sum_{i=1}^{n} W_i R_{ij} = \sum_{i=1}^{n} \frac{R_i (S - |R_i - \frac{1}{n} \sum_{j=1}^{m} R_{ij}|)}{\sum_{i=1}^{n} (S - |R_i - \frac{1}{n} \sum_{j=1}^{m} R_{ij}|)}
\]

3.3 Collaborative Evaluation Based on Overall Agreement (CE_AGREE)

In order to further illustrate the advantages of collaborative evaluation method, the proposed CE method also made a further extension to make it more suitable for a variety of applications situation, such as evaluation with initial weights for evaluators. In the part, collaborative evaluation based on overall agreement (CE_AGREE) is proposed to get valuable evaluation information, such as influential degree of the evaluators to objects evaluated. The influence information shows that positive value indicates the size of the impact on the single evaluator to group evaluation, and negative value indicates that the evaluator is counterproductive to group evaluation. CE_AGREE method can be computed as following steps.

Step 1: For any evaluated object \( O_j \), calculate group score for each evaluated object, shown as follows:

\[
\bar{R}_{ij} = \sum_{i=1}^{n} u_i R_{ij} \quad (j=1, 2, ..., m)
\]

where \( u_i \) denotes the initial weights for each evaluator. When all evaluators have same weights, it calculates average score for each evaluated object. In the following example, \( u_i=1/n \), where \( n \) is overall number of evaluators.
Step 2: For any evaluated object \( O_j \), calculate the difference between group score and rating score given by evaluator \( U_i \) as follows:

\[
V_{ij} = |R_{ij} - \overline{R}_{ij}| = |R_{ij} - \frac{\sum_{t=1}^{n} u_t R_{ij}}{|S|}| \quad (i=1,2,...,n; \quad j=1,2,...,m)
\]

Step 3: calculate the weights of the evaluator to object based on evaluating difference and the overall agreement measure [Chen 2012; José et al. 2007].

Firstly, calculate the consensus level \( C_{O_i} \) of evaluated object \( O_j \), shown as follows:

\[
C_{O_i} = \sum_{j=1}^{m} (1 - \frac{|R_{ij} - \overline{R}_{ij}|}{|S|}) u_j = \frac{\sum_{j=1}^{m} (1 - \frac{|R_{ij} - \overline{R}_{ij}|}{|S|}) u_j}{|S|}
\]

where \( S \) denotes evaluation scale, \( 1 \leq i \leq n \), and \( 1 \leq j \leq m \).

Secondly, calculate the consensus level \( C_{O_i} \) of evaluated object \( O_j \) without evaluator \( U_z \), shown as follows:

\[
C_{O_i} = \sum_{k \in U \setminus \{z\}} (1 - \frac{|R_{ij} - \overline{R}_{ij}|}{|S|}) \beta_k = \sum_{k \in U \setminus \{z\}} (1 - \frac{|R_{ij} - \overline{R}_{ij}|}{|S|}) \beta_k
\]

where \( U = \{1,2,...,n\} \); \( 1 \leq i \leq n \), \( 1 \leq j \leq m \); and \( k \in U \setminus \{z\} \) denotes \( k \in U \) and \( k \neq z \); \( \beta_k = u_k \sum_{i \in U \setminus \{z\}} u_i \); where \( i \in U \setminus \{z\} \) denotes \( i \in U \) and \( i \neq z \);

Thirdly, calculate the contribution \( D_{ij} \) of evaluator \( U_i \) on evaluated object \( O_j \), shown as follows:

\[
D_{ij} = C_{O_j} - C_{O_{ij}}
\]

The larger the value of \( D_{ij} \), the higher the contribution (influential degree) of evaluator \( U_i \) to evaluated object \( O_j \), where \( 1 \leq i \leq n \), \( 1 \leq j \leq m \).

Finally, update the weights of the evaluator to evaluated object, shown as follows:

\[
W_{ij}' = u_i + D_{ij}
\]

where \( 1 \leq i \leq n \), \( 1 \leq j \leq m \).

Step 4: Normalize the evaluating weight, shown as follows:

\[
W_{ij} = \frac{W_{ij}'}{\sum_{i=1}^{n} W_{ij}'}
\]

Where \( 1 \leq i \leq n \), \( 1 \leq j \leq m \).

Step 5: Computing final evaluation score, shown as follows:

\[
R_{ij} = \sum_{j=1}^{m} W_{ij} R_{ij}
\]

Where \( 1 \leq i \leq n \), \( 1 \leq j \leq m \).

4 EXPERIMENTS

In this section, experiments are conducted to verify the proposed methods. In our experiments two datasets are used. The first dataset is a real rating data collected from students in Project Management
course in a university. Its task is to evaluate the presentation of students in the course. There are 16 students who can evaluate each other. Their scores are shown in Table 2. Each student marked for others, for example, $(S_1, S_2) = 86$ means that student 1 marked student 2 as 86.

<table>
<thead>
<tr>
<th>Student</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>89.31</td>
<td>86.88</td>
<td>87.50</td>
<td>86.56</td>
<td>89.81</td>
<td>85.69</td>
<td>85.94</td>
<td>90.19</td>
</tr>
<tr>
<td>AER</td>
<td>6</td>
<td>10</td>
<td>8</td>
<td>13</td>
<td>4</td>
<td>16</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>CE_DIFFS</td>
<td>89.33</td>
<td>86.88</td>
<td>87.50</td>
<td>86.58</td>
<td>89.82</td>
<td>85.72</td>
<td>85.94</td>
<td>90.17</td>
</tr>
<tr>
<td>CE_DIFFR</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>13</td>
<td>4</td>
<td>16</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>CE_AGREES</td>
<td>89.33</td>
<td>86.88</td>
<td>87.50</td>
<td>86.58</td>
<td>89.82</td>
<td>85.72</td>
<td>85.94</td>
<td>90.17</td>
</tr>
<tr>
<td>CE_AGREE</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>13</td>
<td>4</td>
<td>16</td>
<td>15</td>
<td>2</td>
</tr>
</tbody>
</table>

In our experiment using the first data set, CE_DIFF method, CE_AGREE method and averaging method were compared. As shown in Table 3, AES, CE_DIFFS and CE_AGREES respectively means the score computed by averaging method, CE_DIFF method and CE_AGREE method, while AER, CE_DIFFR and CE_AGREE respectively means the rank computed by averaging method, CE_DIFF method and CE_AGREE method. We can see that each student has same rank and nearly same score for averaging method, CE_DIFF method and CE_AGREE method from Table 3. However it is key to find the irregular evaluation for evaluation method when some evaluators give too low or too high rating to the evaluator.

<table>
<thead>
<tr>
<th>Student</th>
<th>S9</th>
<th>S10</th>
<th>S11</th>
<th>S12</th>
<th>S13</th>
<th>S14</th>
<th>S15</th>
<th>S16</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>88.94</td>
<td>86.88</td>
<td>89.63</td>
<td>86.00</td>
<td>89.94</td>
<td>87.00</td>
<td>86.81</td>
<td>90.63</td>
</tr>
<tr>
<td>AER</td>
<td>7</td>
<td>11</td>
<td>5</td>
<td>14</td>
<td>3</td>
<td>9</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>CE_DIFFS</td>
<td>88.94</td>
<td>86.91</td>
<td>89.64</td>
<td>86.01</td>
<td>89.94</td>
<td>87.02</td>
<td>86.82</td>
<td>90.63</td>
</tr>
<tr>
<td>CE_DIFFR</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>14</td>
<td>3</td>
<td>9</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>CE_AGREES</td>
<td>88.94</td>
<td>86.91</td>
<td>89.64</td>
<td>86.01</td>
<td>89.93</td>
<td>87.02</td>
<td>86.82</td>
<td>90.63</td>
</tr>
</tbody>
</table>
Table 3. Results of Methods Used in First Data Set

So in order to further understand the difference of the three methods, one score is controlled. We changed the score $R_{16,16}$ from 100 to 0 step by step where the length of step is 10. Each time the comparison was made. Detail score and rank of the student $S_{16}$ by average method, CE_DIFF method and CE_AGREE method are shown in Figure 1 when $R_{16,16}$ vary from 0 to 100. As shown in Figure 1, average method is a linear method where the rank of $S_{16}$ rises with increasing of $R_{16,16}$ while CE_DIFF method and CE_AGREE method are a nonlinear method. It is shown that CE_DIFF method and CE_AGREE method are more stable than the average method. The variation for CE_DIFF method and CE_AGREE method are smaller than variation for averaging method.

![Score for S16 with variation of R16,16](image1)

**a. The variation of score for $S_{16}$ with variation of $R_{16,16}$**

![Rank for S16 with variation of R16,16](image2)

**b. The variation of rank for $S_{16}$ with variation of $R_{16,16}$**

*Figure 1. The Results of Experiment Used in First Data Set*

The second dataset is a randomly generated rating dataset with 100 evaluators and 100 evaluated objects. Rating value ranges from 60 to 100. With the same experiment using the second data set, CE_DIFF method, CE_AGREE method and averaging method were compared. We randomly selected 20 objects evaluated to show the results in Table 4. The averaging method, CE_DIFF method and CE_AGREE method have different the score and rank.

<table>
<thead>
<tr>
<th>Object</th>
<th>$O_1$</th>
<th>$O_2$</th>
<th>$O_3$</th>
<th>$O_4$</th>
<th>$O_5$</th>
<th>$O_6$</th>
<th>$O_7$</th>
<th>$O_8$</th>
<th>$O_9$</th>
<th>$O_{10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>79.49</td>
<td>78.06</td>
<td>80.34</td>
<td>81.9</td>
<td>79.58</td>
<td>79.52</td>
<td>79.97</td>
<td>79.45</td>
<td>81.07</td>
<td>79.26</td>
</tr>
<tr>
<td>AER</td>
<td>58</td>
<td>92</td>
<td>32</td>
<td>2</td>
<td>53</td>
<td>56</td>
<td>43</td>
<td>60</td>
<td>9</td>
<td>64</td>
</tr>
<tr>
<td>CE_DIFFS</td>
<td>79.41</td>
<td>78.01</td>
<td>80.34</td>
<td>81.95</td>
<td>79.6</td>
<td>79.49</td>
<td>79.98</td>
<td>79.5</td>
<td>81.17</td>
<td>79.31</td>
</tr>
<tr>
<td>CE_DIFFR</td>
<td>60</td>
<td>91</td>
<td>35</td>
<td>2</td>
<td>53</td>
<td>57</td>
<td>43</td>
<td>56</td>
<td>9</td>
<td>63</td>
</tr>
<tr>
<td>CE_AGREES</td>
<td>79.42</td>
<td>78.02</td>
<td>80.34</td>
<td>81.94</td>
<td>79.6</td>
<td>79.49</td>
<td>79.98</td>
<td>79.5</td>
<td>81.17</td>
<td>79.3</td>
</tr>
<tr>
<td>CE_AGREE</td>
<td>60</td>
<td>92</td>
<td>34</td>
<td>2</td>
<td>53</td>
<td>57</td>
<td>43</td>
<td>56</td>
<td>9</td>
<td>63</td>
</tr>
</tbody>
</table>
Table 4. Results of Methods Used in Second Data Set

<table>
<thead>
<tr>
<th>Object</th>
<th>O₁₁</th>
<th>O₁₂</th>
<th>O₁₃</th>
<th>O₁₄</th>
<th>O₁₅</th>
<th>O₁₆</th>
<th>O₁₇</th>
<th>O₁₈</th>
<th>O₁₉</th>
<th>O₂₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>AES</td>
<td>79.47</td>
<td>81.28</td>
<td>79.32</td>
<td>79.03</td>
<td>78.91</td>
<td>78.94</td>
<td>79.85</td>
<td>78.98</td>
<td>81.16</td>
<td>80.39</td>
</tr>
<tr>
<td>AER</td>
<td>59</td>
<td>7</td>
<td>63</td>
<td>67</td>
<td>72</td>
<td>70</td>
<td>48</td>
<td>69</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>CE_DIFFS</td>
<td>79.43</td>
<td>81.36</td>
<td>79.3</td>
<td>78.98</td>
<td>78.88</td>
<td>78.95</td>
<td>79.82</td>
<td>78.99</td>
<td>81.26</td>
<td>80.45</td>
</tr>
<tr>
<td>CE_DIFFR</td>
<td>59</td>
<td>7</td>
<td>64</td>
<td>68</td>
<td>71</td>
<td>70</td>
<td>49</td>
<td>67</td>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>CE_AGREES</td>
<td>79.43</td>
<td>81.36</td>
<td>79.3</td>
<td>78.99</td>
<td>78.88</td>
<td>78.95</td>
<td>79.82</td>
<td>78.99</td>
<td>81.25</td>
<td>80.45</td>
</tr>
<tr>
<td>CE_AGREE</td>
<td>59</td>
<td>7</td>
<td>64</td>
<td>68</td>
<td>71</td>
<td>70</td>
<td>49</td>
<td>67</td>
<td>8</td>
<td>31</td>
</tr>
</tbody>
</table>

Then in order to further understand that our proposed methods can offset the effect resulted by some evaluator’s irregular evaluation, one score is controlled. We changed the score \( R_{100,100} \) in the second data set from 100 to 0 step by step where the length of step is 10. Detail score and rank of the object 100 by average method, CE_DIFF method and CE_AGREE method are shown in Figure 2 when \( R_{100,100} \) vary from 0 to 100. There is same conclusions of the first data set.

![Score for O₁₀₀](image1)

**a. The variation of score for O₁₀₀ with variation of R₁₀₀,₁₀₀**

![Rank for O₁₀₀](image2)

**b. The variation of rank for O₁₀₀ with variation of R₁₀₀,₁₀₀**

*Figure 2. The Results of Experiment Used in Second Data Set*

As shown in Figure 2(a), average method is a linear method where the score of O₁₀₀ rise with increasing of \( R_{100,100} \) while CE_DIFF method and CE_AGREE method are a nonlinear method. It is shown that CE_DIFF method and CE_AGREE method are more stable than the average method. In Figure 2(b) the rank variation of O₁₀₀ for CE_DIFF method and CE_AGREE method are smaller than rank variation of O₁₀₀ for averaging method. Furthermore, the influential degree of active evaluator by CE_AGREE method can be obtained when rating of active evaluator vary from 0 to 100. And it is shown that the larger the value of D regardless of positive or negative, the higher of difference between the active evaluator and the other evaluators to identical objects evaluated. As a result, this value can be adjusted to the weight of each evaluator for objects evaluated, which is more in line with the actual requirements, while these results cannot be obtained by averaging method. Moreover, some
inequitable or unprofessional evaluators will be found out in some application situations by CE_AGREE method.

5 CONCLUSION

Evaluation is one of basic and difficult research issues in management science. Inspired by collaboration embodied in Web2.0, in this paper the author creatively applied collaboration idea to management evaluation, and a collaborative evaluation theory was proposed. The advantages, disadvantages, the scope and steps of the application were detailed.

Generally, traditional group evaluation method based on rating preference did not consider the difference of each individual evaluator and could aggregate group result through averaging the scores obtained by each objects evaluated. In this work, we add a weighting factor to the traditional group evaluation method by collaboratively finding the irregular evaluation and assigning the different weight to each evaluator. We expect that this weighting factor can offset the effect resulted by some evaluator’s irregular evaluation through considering the contribution of the individual evaluation to the collective evaluation. To verify the effectiveness of our proposed two collaborative methods, experiments are performed using an active example and a synthetic data set. The results indicate that the proposed methods are more stable than traditional group evaluation method and demonstrate that our approach can effectively weaken the irregular evaluations.

In the future, research on method designing of CE for different application background will be implemented. Furthermore, research on applying collaborative evaluation to different forms of individual evaluation will be extended.

ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundation of China(No.71271209), Beijing Natural Science Foundation(No.4132067), Humanity and Social Science Youth Foundation of Ministry of Education of China(No.11YJC630268), the Opening Project of State Key Laboratory of Digital Publishing Technology, the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China.

References


