

Reducing Positive Leniency in Fuzzy Measure Ratings

Ting-Yu Chen¹ Jih-Chang Wang²

¹ Department of Business Administration, Chang Gung University

² Department of Information Management, Chang Gung University

Abstract

Fuzzy measures have been widely used to determine the degrees of subjective importance of evaluation items. However, the leniency error may exist when most attributes are assigned unduly high ratings. Because respondents often assign similarly complimentary scores, errors of positive leniency make it difficult to differentiate the importance of decision attributes. To reduce positive leniency in fuzzy measure ratings, we develop a method by comparison of fuzzy number-valued fuzzy measures using a fuzzy distance measure.

Keywords: Fuzzy Measure, Positive Leniency, Fuzzy Distance Measure.

1. Introduction

The past research consistently reveals that humans have tremendous limitations in processing information (Noe, 2003). People sometimes make some errors called “error of leniency”, “error of venter tendency”, and “halo effect” when they evaluate something or make decisions. Error of leniency that appears often in the subjective measures of performance is usually found in the human resource and decision-making areas (Bass, 1956).

The bias of leniency is one of the distributional errors and it is often found in fuzzy measure ratings. A fuzzy measure, introduced by Sugeno (1974), is a subjective scale for the degrees of fuzziness. Fuzzy measures are suitable in analyzing human subjective evaluation processes. When fuzzy measures are used to real world problems, the candidate sets generally correspond to the evaluation items and the values of fuzzy measures stand for the grades of importance of corresponding evaluation items. Fuzzy measures have been so far used to determine the degrees of subjective importance of evaluation items in numerous studies. In this study, we also use fuzzy measures to express and deal with human subjective importance of evaluation items for multiple attribute decision-making analysis.

When fuzzy measures are used to determine attribute importance, most respondents are assigned very important evaluations; few or none are assigned

unfavorable scores. It is difficult to distinguish grades of importance among attributes rated by the same person when the positive leniency occurred. Glickman (1955) has noted the serious effects of the lenient merit rating on the motivation of the ratee. Leniency makes it difficult to differentiate person’s merit, because ratees are assigned similarly complimentary scores. Leniency error is very usual in human’s life and researches of it have been ongoing for more than 50 years. Freeman and Taylor (1950) believed leniency is the most glaring error that raters make. In this study, we also want to develop a method to reduce the leniency error in fuzzy measure rating.

2. Approaches to Reduce Leniency

(1) Changing the number of response categories

Some studies were hoped to reform the lenient bias by designing a new evaluating scale. Bass (1956) introduced a discriminate binary scoring eliminates the leniency effect in these ratings without loss of internal consistency or validity. Blanz introduced a new merit rating method in 1965. The mixed standard scale (MSS), was developed in Finland and applied there in several different occupations at the worker level. Blanz and Ghiselli (1972) studied the supervisory ratings obtained on 100 “middle managers” in a chemical and aluminum plant in the United States. Saal and Landy (1977) made a comparison between the mixed standard scale and another conventional rating scale, they found the MSS would be characterized by less leniency. But Dickinson and Glebocki (1990) found no significant differences between the MSS formats in leniency and halo effect.

(2) Experiment manipulations

Holzbach (1978) defined the self-appraisal leniency as an overestimation of self-reported performance as better than the observed performance. Farh and Werbel (1986) examined the factors that may affect the leniency bias of self-appraisals. They designed a 2×2 factorial experiment to assess the effect of the purpose of the appraisal and expectation of validation on the amount of leniency. Harris et al. (1995) examined 223 ratees by administrative-based performance appraisal and research-based performance appraisal ratings and found that

administrative ratings were more lenient than research-based ratings.

(3) Training approach

Rating accuracy training (RAT) emphasizes the content of various performance dimensions and the trainees should pay more attention to how rates' behavior fit the performance descriptions. Rater error training (RET) attempts to make managers aware of rating errors and helps them develop strategies for minimizing those errors (Ivancevich, 1979). But Bernardin and Pence (1980) found the RET reduced the rating errors can also reduce accuracy.

3. Research Method

Many scholars used the scales designing, experiment manipulating, training to reduce the leniency. Training is the most often used in reducing leniency, but it is not suit for our study. We are not able to train the respondents before we start the study, so we develop a new method to see if it can reduce the leniency effect.

A fuzzy number-valued fuzzy measure, proposed by Zhang (1992), is used to identify the grade of importance for evaluation attributes. Let F^* be the set of all fuzzy numbers and $A \in F^*$. An α -level interval of fuzzy number A is denoted as $[\alpha A^-, \alpha A^+]$. Then, by the decomposition theorem of a fuzzy set, $A = \cup \alpha [\alpha A^-, \alpha A^+]$ for every $A \in F^*$. In addition, let $F_+^* = \{A; A \geq 0, A \in F^*\}$. A fuzzy number-valued fuzzy measure (i.e., a (z)fuzzy measure) on $F^*(X)$ is a fuzzy number-valued fuzzy set function $\mu: F^*(X) \rightarrow F_+^*$. We apply (z)fuzzy measures to determine the salience of decision attributes.

If attribute salience is interpreted as a linguistic variable, then its term set T represents the scale position marked by respondents. T is a mapping from X to $[0,1]$ (i.e., $T: X \rightarrow [0,1]$), where X is a nonempty attribute set and $X = \{x_1, x_2, \dots, x_n\}$. Note that $T(x_i)$ for a single element x_i is denoted by T^i . In this study, we conduct a reasonable expression of attribute importance by TFNs. The verbal term set used in 6-scale is {very unimportant, unimportant, fairly unimportant, fairly important, important, very important}. Each of the basic linguistic terms is assigned one of six TFNs by a semantic rule. The input data contain the respondent's subjective judgment on attribute importance expressed by a linguistic term T^i .

A fuzzy distance of fuzzy numbers A and B is denoted as $D(A,B)$, where

$$D(A,B) = \bigcup_{\alpha \in [0,1]} \alpha \left| \left[\alpha A^- - \alpha B^- \right] \right|$$

$$\sup_{\alpha \leq \eta \leq 1} \left| \eta A^- - \eta B^- \right| \vee \left| \eta A^+ - \eta B^+ \right|$$

for $A, B \in F^*$. Tran and Duckstein (2002) introduced a new approach for ranking fuzzy numbers based on a distance measure. In this study, we apply Tran and Duckstein's distance measure for fuzzy numbers to adjust overestimated ratings of attribute importance.

Since TFNs are the most widely-used fuzzy numbers, this study assumes that a fuzzy number-valued fuzzy set function μ on (z)fuzzy measure space $(X, F^*(X), \mu)$ is a TFN for simplicity. Given two TFNs that $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$. Tran and Duckstein (2002) suggested that $w(\alpha)$ is an increasing function so as to assign greater weight to the distance between higher α -level intervals. When $w(\alpha) = 1$, the distances have equal weights at different α levels. When $w(\alpha) = \alpha$, the distances have more weights at higher α levels.

The (z)fuzzy measure of the most important attribute, x_k , is denoted by $\mu(T^k)$ and $\mu(T^i)$'s are the (z)fuzzy measures of other less salient attributes, where $i \neq k$. Let $w(\tau)$ be a weighting function towards the distance between $\mu(T^i)$ and $\mu(T^k)$ at different τ levels, where $\tau \in [0,1]$. When $w(\tau) = 1$, we have:

$$D_{TFN}(\mu(T^i), \mu(T^k), 1) = \frac{1}{225} + (a_{i3} - a_{k2} - 0.2) \\ \times (2a_{i3} - \frac{1}{2}a_{k3} - a_{k2} - \frac{1}{2}a_{k1} - 0.4) \\ + \frac{1}{9} \left[(a_{k3} - a_{k2})^2 + (a_{k2} - a_{k1})^2 \right. \\ \left. - (a_{k3} - a_{k2})(a_{k2} - a_{k1}) \right]$$

When $w(\tau) = \tau$, we have:

$$D_{TFN}(\mu(T^i), \mu(T^k), \tau) = \frac{1}{450} + \frac{1}{3}(a_{i3} - a_{k2} - 0.2) \\ \times (5a_{i3} - a_{k3} - 3a_{k2} - a_{k1} - 1) \\ + \frac{1}{18} \left[(a_{k3} - a_{k2})^2 + (a_{k2} - a_{k1})^2 \right. \\ \left. - (a_{k3} - a_{k2})(a_{k2} - a_{k1}) \right]$$

Given a_{k1}, a_{k2}, a_{k3} , and D_{TFN} , we can derive the a_{i3} value by above equations. The approach supposes that the fuzzy numbers of all decision attributes have the same spread for convenience. Note that the solution of a_{i3} has to be smaller than a_{k3} because the grade of importance, (a_{k1}, a_{k2}, a_{k3}) , is highest. The above approach has its exceptions in the cases of $a_{i3} < 0.4$. To obtain a meaningful fuzzy number, a_{ij} must not be smaller than zero for $j=1,2,3$. Thus,

$$\begin{cases} a_{i1} = \max(a_{i3} - 0.4, 0), \\ a_{i2} = \max(a_{i3} - 0.2, 0). \end{cases}$$

We use the above-mentioned approach to characterize the grade of importance associating with each decision attribute, exclusive of the most important one. The input data contain the respondent's subjective judgment on the most important attribute x_k

and its corresponding linguistic term T^k , and relative importance distance of other attributes to x_k .

The flowchart of our proposed algorithm is indicated in Figure 1.

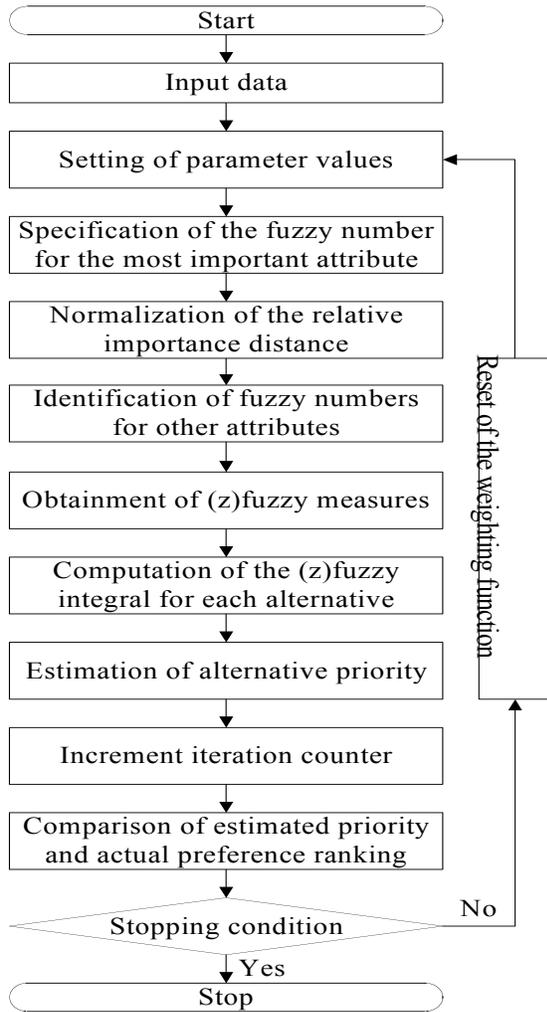


Figure 1: Flowchart of the importance-assessing algorithm.

The importance-assessing method given here is chiefly concerned with an individual's judgment on the relative value of attributes. Therefore, an outcome-oriented approach is necessary to advance the estimation accuracy so as to obtain a more elegant method.

4. Empirical Study

The empirical study intends to inquire into consumer choice problems to examine the estimation results of different importance-assessing methods. Three kinds of consumer goods regularly used by Taiwanese college students were selected from the 2005 E-ICP (Eastern integrated consumer profile) database: convenience goods, electronic goods, and health foods.

Three product categories are taken as empirical cases, including pudding in convenience goods, CD burner in electronic goods, and health tea in health foods. For each product category, we conducted a face-to-face interview on experienced college students in Taoyuan County about the data items. The sample sizes of three product categories were designated as 40, respectively.

We adopt the unadjusted (z)fuzzy measures (when $w(\tau)=1$ or $w(\tau)=\tau$) to draw a comparison. Moreover, the values of (z)fuzzy measures and typical fuzzy measures are supposed to satisfy the finite λ -rule. The main input data for typical fuzzy measures (FMs) are linguistic values about importance of each attribute. As for the eigenvector method, the DM is supposed to judge the relative importance of two attributes, and thus the scaling ratios in the pairwise comparison matrix are the main input data. The weighted least square (WLS) method is also applied to Saaty's comparison matrix.

There are five kinds of performance indices for outcome prediction comparison, including total order index (I_1), partial order index (I_2), average rank correlation coefficient (I_3), average residual sum (I_4), and average error sum of squares (I_5). The first three indices are developed for the ordinal preferences of alternatives given. The last two indices treat cardinal preferences of alternatives.

Table 1 summarizes the major results of performance indices using four importance-assessing methods. In view of the total order index (I_1), our proposed method performs better than other methods in pudding and CD burner categories; while the weighted least square method performs best in the health tea category. In regard to the partial order index (I_2), the accuracy level of our proposed method is highest in CD burner and health tea categories. In addition, the superior methods are (z)FM and FM in the pudding category. Concerning average rank correlation coefficient (I_3), the best methods are (z)FM and FM in the pudding category, (z)FM in the CD burner category, and WLS in the health tea category. Our method performs well both in the average residual sum (I_4) and average error sum of squares (I_5). We can be fairly certain that the proposed importance-assessing method is more practical for consumer choice problems. The empirical results support the validity of using a distance between (z)fuzzy measures to determine attribute importance.

The observation of λ -value changes between (z)fuzzy measures and typical fuzzy measures will be implemented as a method for eliminating errors of positive leniency in rating systems. The comparison results are presented in Table 2 ($w(\alpha)=1$). As sketched here, most positive leniency errors can be improved.

Table 1: Results of performance indices using four methods.

Product category y	Performance index	(z)FM ¹		FM ²	AHP ³	WLS ⁴
		w(τ)=1	w(τ)=τ			
Pudding	I ₁	0.9743	0.9487	0.9487	0.8717	0.8717
	I ₂	0.9332	0.9332	0.9332	0.9167	0.9167
	I ₃	0.9167	0.8654	0.9167	0.8910	0.8910
	I ₄	0.0296	0.0301	0.0315	0.0318	0.0323
	I ₅	0.0074	0.0073	0.0083	0.0092	0.0092
CD burner	I ₁	0.8461	0.8205	0.7948	0.6923	0.6923
	I ₂	0.8672	0.8587	0.8631	0.7736	0.8077
	I ₃	0.7372	0.7244	0.6538	0.5513	0.6282
	I ₄	0.0380	0.0377	0.0419	0.0410	0.0412
	I ₅	0.0029	0.0029	0.0033	0.0031	0.0110
Health tea	I ₁	0.7750	0.7750	0.7750	0.7750	0.8250
	I ₂	0.9168	0.9125	0.9043	0.8960	0.9128
	I ₃	0.8313	0.8250	0.7688	0.8188	0.8688
	I ₄	0.0322	0.0330	0.0362	0.0350	0.0345
	I ₅	0.0023	0.0026	0.0032	0.0026	0.0025

Table 2: Improvement results of errors of positive leniency.

Product category	α _{level}	Left side			Right side		
		+	O	-	+	O	-
Pudding	0.0	40 ¹ (100 ²)	0 (0.0)	0 (0.0)	0 (0.0)	26 (65.0)	14 (35.0)
	0.2	40 (100)	0 (0.0)	0 (0.0)	0 (0.0)	26 (65.0)	14 (35.0)
	0.4	40 (100)	0 (0.0)	0 (0.0)	5 (12.5)	27 (67.5)	8 (20.0)
	0.6	40 (100)	0 (0.0)	0 (0.0)	10 (25.0)	26 (65.0)	4 (10.0)
	0.8	40 (100)	0 (0.0)	0 (0.0)	12 (30.0)	26 (65.0)	2 (5.0)
	1.0	14 (35.0)	26 (65.0)	0 (0.0)	14 (35.0)	26 (65.0)	0 (0.0)
CD burner	0.0	40 (100)	0 (0.0)	0 (0.0)	1 (2.5)	22 (55.0)	17 (42.5)
	0.2	40 (100)	0 (0.0)	0 (0.0)	4 (10.0)	22 (55.0)	14 (35.0)
	0.4	40 (100)	0 (0.0)	0 (0.0)	7 (17.5)	22 (55.0)	11 (27.5)
	0.6	40 (100)	0 (0.0)	0 (0.0)	13 (32.5)	22 (55.0)	5 (12.5)
	0.8	40 (100)	0 (0.0)	0 (0.0)	16 (40.0)	22 (55.0)	2 (5.0)
	1.0	18 (45.0)	22 (55.0)	0 (0.0)	18 (45.0)	22 (55.0)	0 (0.0)
Healthy tea	0.0	40 (100)	0 (0.0)	0 (0.0)	1 (2.5)	13 (32.5)	16 (40.0)
	0.2	40 (100)	0 (0.0)	0 (0.0)	1 (2.5)	13 (32.5)	16 (40.0)
	0.4	40 (100)	0 (0.0)	0 (0.0)	7 (17.5)	13 (32.5)	10 (25.0)
	0.6	40 (100)	0 (0.0)	0 (0.0)	11 (27.5)	23 (57.5)	6 (15.0)
	0.8	40 (100)	0 (0.0)	0 (0.0)	16 (40.0)	13 (32.5)	1 (2.5)
	1.0	17 (42.5)	23 (57.5)	0 (0.0)	17 (42.5)	23 (57.5)	0 (0.0)

5. Conclusions

Based on the distance between (z)fuzzy measures, we developed a new importance-assessing method to determine the salience of decision attributes. From an outcome-oriented point of view, several empirical consumer decision-making problems were addressed to examine the applicability of the current method in contrast to several popular methods for assessing attribute importance. The comparison results show that the proposed method makes a relatively correct prediction of decision outcomes than several existing importance-assessing approaches. Furthermore, our method deals with the matter of positive leniency and helps overcome the problem of the indiscriminate weights. Meanwhile, for each numerical case, there are only n input data required by our proposed method, and this can reduce the difficulty of data collection in

practice. The easy-to-use and easy-to-understand characteristics of our new approach make it valuable to multiattribute decision making reality.

6. References

- [1] Bass, B.M. (1956), "Reducing leniency in merit ratings," *Personnel Psychology* 9, 359-369.
- [2] Bernardin, H.J., and Pence, E.C., (1980), "The effects of rater training: Creating new response sets and decreasing accuracy," *Journal of Applied Psychology* 65, 60-66.
- [3] Blanz, F., (1965), "A new merit rating method," Helsinki: Unpublished doctoral dissertation.
- [4] Blanz, F., and Ghiselli, E.E., (1972), "The mixed standard scale: A new rating system," *Personnel Psychology* 25, 185-199.
- [5] Dickinson, T.L., and Glebocki, G.G., (1990), "Modifications in the format of the mixed standard scale," *Organizational Behavior and Human Decision Processes* 47, 124-137.
- [6] Freeman, G.L., and Taylor, E.K. (1950), *How to Pick Leaders*, New York: Funk and Wagnalls.
- [7] Glickman, A.S. (1955), "Effects of negatively skewed ratings on motivations of the rated," *Personnel Psychology* 8, 39-47.
- [8] Harris, M.M., Smith, D.E., and Champagne, D., (1995), "A field of performance appraisal purpose: Research- versus administrative-based ratings," *Personnel Psychology* 48, 1, 151-160.
- [9] Holzbach, R.L., (1978), "Rater bias in performance ratings: Superior, self-, and peer ratings," *Journal of Applied Psychology* 63, 579-588.
- [10] Ivancevich, J.M., (1979), "Longitudinal study of the effects of rater training on psychometric error in ratings," *Journal of Applied Psychology* 64, 502-508.
- [11] Noe, R.A., (2003), *Human Resource Management*, 4th ed, McGraw Hill, New York, 358.
- [12] Saal, F.E., and Landy, F.J., (1977), "The mixed standard rating scale: An evaluation," *Organizational behavior and human performance* 18, 19-35.
- [13] Sugeno, M. (1974), *Theory of Fuzzy Integrals and its Applications*, Ph.D. Dissertation. Tokyo Institute of Technology.
- [14] Tran, L., and Duckstein, L. (2002), "Comparison of fuzzy numbers using a fuzzy distance measure," *Fuzzy Sets and Systems* 130, 331-341.
- [15] Zhang, G.-Q. (1992), "On fuzzy number-valued fuzzy measures defined by fuzzy number-valued fuzzy integrals I," *Fuzzy Sets and Systems* 45, 227-237.