Comparison of Weights in TOPSIS Models

D.L. Olson
Department of Management, CBA
University of Nebraska
Lincoln, NE 68506, U.S.A.
Dolson3@unl.edu

Abstract—TOPSIS is a multiple criteria method to identify solutions from a finite set of alternatives based upon simultaneous minimization of distance from an ideal point and maximization of distance from a nadir point. TOPSIS can incorporate relative weights of criterion importance. This paper reviews several applications of TOPSIS using different weighting schemes and different distance metrics, and compares results of different sets of weights applied to a previously used set of multiple criteria data. Comparison is also made against SMART and centroid weighting schemes. TOPSIS was not found to be more accurate, but was quite close in accuracy. Using first-order and second-order metrics were found to be quite good, but the infinite order (Tchebycheff norm, L-∞) was found to decline in accuracy. © 2004 Elsevier Science Ltd. All rights reserved.

1. INTRODUCTION

There are a variety of multiple criteria techniques to aid selection in conditions of multiple criteria. The acronym TOPSIS stands for technique for preference by similarity to the ideal solution [1]. TOPSIS was initially presented by Hwang and Yoon [2], Lai *et al.* [3], and Yoon and Hwang [4]. TOPSIS is attractive in that limited subjective input is needed from decision makers. The only subjective input needed is weights. In this study, we consider three different weight sources, reflected by equal weights, centroid weights, and weights obtained by regression. All three sources are linear. We also consider three commonly used metrics (L_1, L_2, and L_∞, or alternative names P1, P2, and Pinfinity) to measure distance from ideal or nadir points.

TOPSIS has been applied to a number of applications [5,6], although it is not nearly as widely applied as other multiattribute methods [7]. A variant of TOPSIS was used for selection of grippers in flexible manufacturing [8,9]. TOPSIS was applied to financial investment in advanced manufacturing systems [10]. In other manufacturing applications, it has been used in a case selecting a manufacturing process [11] and in an application selecting robotic processes [12–14]. Neural network approaches to obtain weights for TOPSIS have been applied [10], and fuzzy set extensions implemented [15,16]. TOPSIS has also been used to compare company performances [17] and financial ratio performance within a specific industry [18].

2. TOPSIS METHOD

The idea of TOPSIS can be expressed in a series of steps.

(1) Obtain performance data for n alternatives over k criteria. Raw measurements are usually standardized, converting raw measures x_{ij} into standardized measures s_{ij}.

(2) Develop a set of importance weights w_k, for each of the criteria. The basis for these weights can be anything, but, usually, is *ad hoc* reflective of relative importance. Scale is not an issue if standardizing was accomplished in Step 1.

(3) Identify the ideal alternative (extreme performance on each criterion) s^+.

(4) Identify the nadir alternative (reverse extreme performance on each criterion) s^-.

(5) Develop a distance measure over each criterion to both ideal (D^+) and nadir (D^-).

(6) For each alternative, determine a ratio R equal to the distance to the nadir divided by the sum of the distance to the nadir and the distance to the ideal,

$$R = \frac{D^-}{D^- + D^+}$$

(7) Rank order alternatives by maximizing the ratio in Step 6.

Thus, TOPSIS minimizes the distance to the ideal alternative while maximizing the distance to the nadir. There are a number of specific procedures that can be used for Step 2 (developing weights), and for Step 5 (distance measures). Additionally, different conventions can be applied to defining best performance (Step 3) and worst performance (Step 4).

A number of distance metrics can be applied. Traditional TOPSIS applied the Euclidean norm (minimization of square root of the sum of squared distances) to ideal and nadir solutions, a second power metric (**P2**). TOPSIS2 is a variant where distance was measured in least absolute value terms, a first power metric (**P1**). Another commonly used metric is the Tchebychev metric, where the minimum maximum difference is the basis for selection. This coincides with an infinite power-term (**P∞**).

A relative advantage of TOPSIS is the ability to identify the best alternative quickly [12]. TOPSIS has been comparatively tested with a number of other multiattribute methods [7]. The other methods primarily focused on generating weights (Step 2 in the prior description), with one method (ELECTRE, [19,20]) including a different way to combine weights and distance measures. TOPSIS was found to perform almost as well as multiplicative additive weights and better than analytic hierarchy process (AHP, [21]) in matching a base prediction model. When there were few criteria, TOPSIS had proportionately more rank reversals. When there were many criteria, TOPSIS differed more from simple additive weight results, and TOPSIS was also affected more with diverse sets of weights. TOPSIS performed less accurately than AHP on both selecting the top ranked alternative and in matching all ranks in this set of simulations.

3. EXPERIMENT

This study tests TOPSIS on baseball data, which was previously used to compare multiple criteria methods [22]. The idea is that performance of a baseball team depends on a number of criteria (ability to hit, power, speed, fielding, and pitching), and baseball provides many cases where these data are available with known ranked outcomes. Here, they are used to compare the effectiveness of alternative sets of weights as well as alternative distance metrics. The alternative sets of weights are developed in three ways: equal weights, weights generated by ordinal rank, and weights generated by regression. The simple multiattribute rating technique (SMART, [23]) uses a systematic procedure to estimate linear weights, much as regression fits data with a linear function. Centroid weights [24,25] utilize ordinal input to generate weights. Alternative distance metrics used are least absolute value, Euclidian, and Tchebycheff weights minimizing the maximum criterion distance.

Tests were run on ten decades of results. Using data mining techniques, the first five years of each decade were used to train (identify relative weights through regression, determine the minimum and maximum performance levels for performance scores). These parameters were then applied to the last five years of data for each decade as a test. Nine combinations of weight generation and distance metrics were tested. Two measures were used:

(1) selection of the top-ranked alternative (predicting the winner), and
(2) matching rank at the end of the season.

4. RESULTS BASED ON WINNERS

Table 1 gives results in terms of the number of correct winners predicted. Because baseball has undergone significant expansion since 1961, the number of teams per season has varied from four teams per division to ten. The term cases in Table 1 refers to the number of seasons played with that number of teams. The proportion of correct results is biased by the number of alternatives. Therefore, results were split out by number of teams per division. For instance, there were ten competitions over the period 1901–1905 involving eight teams. After expansion in 1961, the number of cases will vary as the number of teams per division changed often. The most complex case was for the decade beginning with 1991. Table 1 reports the test results obtained from training over seasons 1991–1995. Thus the test seasons were 1996 through 2000. In seasons 1996 and 1997, each league had three divisions, two of five teams each, and one of four. This was the case for the American League throughout the period. However, the National League changed to two divisions of five teams each and one six team division for the 1998–2000 seasons. Thus, the 1991 decade has seven cases of four teams, 20 cases of five teams, and three cases of six teams.

Of 144 opportunities, the SMART method was correct in 100, for a proportion of 0.69. Results from [16] for the SMART method (based on regression) and the centroid method (weights gen-

Table 1. Number of winners predicted.

Decade	Teams	Cases	SMART	Cent	TC P1	TC P2	TR P1	TR P2	TE P1	TE P2	T P∞
1901	8	10	7	7	7	6	7	7	7	5	4
1911	8	10	7	8	6	6	6	6	6	6	8
1921	8	10	4	5	4	4	5	3	7	5	3
1931	8	10	7	7	7	8	7	7	7	6	5
1941	8	10	7	7	7	6	7	6	6	8	4
1951	8	10	7	6	6	7	7	7	5	6	3
1961	10	6	6	4	5	5	6	5	4	4	1
1961	6	8	6	6	6	5	6	6	6	5	5
1971	6	12	9	10	9	7	8	8	6	7	6
1971	7	8	6	6	5	5	4	5	4	2	2
1981	6	10	8	3	8	8	8	9	7	7	6
1981	7	10	5	4	6	7	5	5	3	5	4
1991	4	7	5	1	5	4	5	5	3	3	1
1991	5	20	15	11	16	18	15	16	13	14	10
1991	6	3	1	0	2	1	1	1	2	1	1
Totals		**144**	**100**	**85**	**99**	**97**	**97**	**96**	**86**	**84**	**63**

TC: TOPSIS using centroid-generated weights

TR: TOPSIS using regression-generated weights

TE: TOPSIS using equal weights

P1: Linear metric (first power)

P2: Euclidian metric (second power)

P∞: Tchebycheff metric (minimize maximum distance—same for T C, T R, T E)

erated from ordinal data) are given for comparison. The TOPSIS method using centroid weights was almost as good as SMART on this measure, while TOPSIS based on regression was quite close as well. TOPSIS based on equal weights was inferior.

Table 2 recaps Table 1 by number of teams, giving proportion of winners selected.

Table 2. Proportion of winners predicted.

Method	4 Teams	5 Teams	6 Teams	7 Teams	8 Teams	10 Teams	Means
Cases	7	20	33	18	60	6	
SMART	0.714	0.75	0.727	0.611	0.650	1.000	**0.694**
Centroid	0.143	0.55	0.576	0.556	0.667	0.667	**0.590**
TOPSIS centroid P1	0.714	0.80	0.758	0.611	0.617	0.833	**0.688**
TOPSIS centroid P2	0.571	0.90	0.636	0.667	0.617	0.833	**0.674**
TOPSIS regression P1	0.714	0.75	0.697	0.500	0.650	1.000	**0.674**
TOPSIS regression P2	0.714	0.80	0.727	0.556	0.600	0.833	**0.667**
TOPSIS equal P1	0.429	0.65	0.636	0.389	0.633	0.667	**0.597**
TOPSIS equal P2	0.429	0.70	0.606	0.389	0.600	0.667	**0.583**
TOPSIS equal P∞	0.143	0.50	0.545	0.333	0.450	0.167	**0.438**

Of course, it is much easier to predict the winning team when there are four alternatives than when there are ten. However, the number of cases (which, here, is the number of opportunities to be right or wrong) is low for seasons with four or ten teams. The SMART method provided a strong base to compare against. TOPSIS based on centroid measures and a P1 metric did better in those seasons involving five, six, or seven teams, and TOPSIS based on regression and a P1 metric did almost as well. The TOPSIS methods with equal weights (including P∞, where weights are washed out) did much worse.

RESULTS BASED ON RANKINGS

To measure differences in complete rankings, Table 3 gives measures of the sum of absolute value differences between predicted and actual finishes for all teams in each division. Cases here are larger, because they reflect the number of finishing places in a season.

Table 3. Sums of absolute errors.

Decade	Teams	Cases	SMART	Cent	TC P1	TC P2	TC P∞	TR P1	TR P2	TR P∞	TE P1	TE P2	TE P∞
1901	8	80	53	57	62	66	93	52	60	93	64	70	93
1911	8	80	65	61	62	62	93	63	66	93	72	72	93
1921	8	80	46	54	62	68	97	54	60	97	60	64	97
1931	8	80	50	53	54	80	95	52	58	95	70	72	95
1941	8	80	60	60	64	74	153	62	72	153	88	80	153
1951	8	80	59	65	66	72	116	64	70	116	84	84	116
1961	10	60	74	73	81	83	134	74	84	134	88	86	134
1961	6	48	29	29	31	29	71	29	29	71	35	37	71
1971	6	72	37	30	40	58	74	36	46	74	52	44	74
1971	7	56	17	31	32	38	70	42	34	70	64	66	70
1981	6	60	34	92	34	46	56	34	36	56	53	53	56
1981	7	70	64	120	64	72	95	64	70	95	75	71	95
1991	4	28	14	32	10	6	32	14	6	32	20	18	32
1991	5	100	60	90	40	50	108	60	52	108	76	74	108
1991	6	18	12	26	10	14	26	12	12	26	14	16	26

The numbers in Table 3 count the sum of absolute differences between predicted and actual. Table 4 recaps Table 3 by number of teams. Values were derived by the following formula:

$$\frac{\sum |R_i - R_j|}{\text{cases}},$$

where R_i is predicted rank and R_j is actual rank.

Table 4. Average absolute rank differences.

Method	4 Teams	5 Teams	6 Teams	7 Teams	8 Teams	10 Teams	Means
Cases	28	100	198	126	480	60	
SMART	0.500	0.600	0.566	0.643	0.694	1.233	**0.679**
Centroid	1.143	0.900	0.894	1.198	0.729	1.217	**0.880**
TOPSIS centroid P1	0.357	0.400	0.581	0.762	0.771	1.350	**0.718**
TOPSIS centroid P2	0.214	0.500	0.742	0.873	0.879	1.383	**0.824**
TOPSIS regression P1	0.500	0.600	0.561	0.841	0.723	1.233	**0.718**
TOPSIS regression P2	0.214	0.520	0.621	0.825	0.804	1.400	**0.761**
TOPSIS equal P1	0.714	0.760	0.778	1.103	0.913	1.467	**0.922**
TOPSIS equal P2	0.643	0.740	0.758	1.087	0.921	1.433	**0.914**
TOPSIS equal P∞	1.143	1.080	1.146	1.310	1.348	2.233	**1.324**

These average differences show that the P1 metric and P2 metric perform very similarly; both are quite a bit more accurate than the Tchebycheff metric. The centroid and regression methods yield very similar results, again, both doing better than the Tchebycheff metric. Since the output is not expected to be normal, significance tests should be nonparametric. Spearman's rho test was used to evaluate significance. For the mean proportions predicted, using the significance test based upon proportions at the error level of 0.05, the only significant differences are between TOPSIS equal P∞ and every other method. Spearman's rho between even the Tchebycheff models and the other TOPSIS methods yield correlations greater than 0.999, due to the large number of tests.

Following [26], the sum of squared rank differences needed to show significant difference by number of teams (alternatives) was used as the basis for comparison at the 5% level. These limits vary by the number of teams (alternatives) in a season, and are given in Table 5.

Table 6 provides the proportion of competitions that were not significantly different by the Spearman's rho limits given in Table 5.

With four cases, the P1 metric for both regression-based weights and equal weights had no errors. The centroid weights did have errors. For those cases with ten teams, the method using the P1 metric and regression-based weights again was the most accurate. Accuracy increases for the most part with the number of teams, primarily due to the very rigid cutoff limit for small alternative cases. This same feature accounts for the perfect accuracy of two methods when there were only four teams (along with small sample size for that case). The P1 metric is slightly more

Table 5. Sum of squared absolute difference cutoff limits—Spearman's rho.

Alternatives	Count Significant at 5% Error
4	1
5	2
6	5
7	13
8	24
10	59

Source: [22]

Table 6. Proportion not significantly different (5% error rate).

Teams	Cent P1	Cent P2	Reg P1	Reg P2	Eq P1	Eq P2	Tcheb
4	.043	0.57	1.00	0.29	1.00	0.14	0.43
5	0.45	0.35	0.20	0.25	0.05	0.10	0.00
6	0.58	0.39	0.67	0.48	0.42	0.42	0.82
7	0.67	0.67	0.67	0.67	0.39	0.39	0.33
8	0.93	0.87	0.95	0.90	0.85	0.87	0.55
10	0.83	0.67	1.00	0.67	0.67	0.67	0.17
Total	0.72	0.51	0.74	0.65	0.58	0.55	0.49

accurate than the P2 metric. Both are much more accurate than the Tchebycheff metric. Of the weighting approaches, regression-based weights are on average slightly more accurate than centroid-generated weights (as would be expected). Both are much more accurate than equal weights (again, as would be expected).

CONCLUSIONS

The inference drawn here is that TOPSIS is not noticeably more accurate than direct weighting methods such as SMART. However, it is quite close in accuracy except when equal weights were applied. Obviously, the key to accuracy in the TOPSIS method is to obtain accurate weights. Either P1 or P2 metrics seem very good, but as the metric increases to infinity, accuracy seems to decline (at least for the extreme of minimizing maximum distance).

REFERENCES

1. K. Yoon, Systems selection by multiple attribute decision making, Ph.D. Dissertation, Kansas State University, (1980).
2. C.L. Hwang and K. Yoon, *Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, New York, (1981).
3. Y.-J. Lai, T.-Y. Liu and C.-L. Hwang, TOPSIS for MODM, *European Journal of Operational Research* **76** (3), 486–500, (1994).
4. K. Yoon and C.L. Hwang, *Multiple Attribute Decision Making: An Introduction*, Sage, Thousand Oaks, CA, (1995).
5. C.L. Hwang, Y.-J. Lai and T.Y. Liu, A new approach for multiple objective decision making, *Computers & Operations Research* **20**, 889–899, (1993).
6. K.P. Yoon, A reconciliation among discrete compromise solutions, *Journal of the Operational Research Society* **38** (3), 277–286, (1987).
7. S.H. Zanakis, A. Solomon, N. Wishart and S. Dublish, Multi-attribute decision making: A simulation comparison of select methods, *European Journal of Operational Research* **107** (3), 507–529, (1998).
8. V.P. Agrawal, V. Kohli and S. Gupta, Computer aided robot selection: The multiple attribute decision making approach, *International Journal of Production Research* **29** (8), 1629–1644, (1991).
9. V.P. Agrawal, A. Verma and S. Agarwal, Computer-aided evaluation and selection of optimum grippers, *International Journal of Production Research* **30** (11), 2713–2732, (1992).
10. G. Kim, C. Park and K.P. Yoon, Identifying investment opportunities for advanced manufacturing system with comparative-integrated performance measurement, *International Journal of Production Economics* **50**, 23–33, (1997).
11. O.L. Chau and C. Parkan, Selection of a manufacturing process with multiple attributes: A case study, *Journal of Engineering Technology Management* **12**, 219–237, (1995).
12. C. Parkan and M.L. Wu, On the equivalence of operational performance measurement and multiple attribute decision making, *International Journal of Production Research* **35** (11), 2963–2988, (1997).
13. C. Parkan and M.L. Wu, Process selection with multiple objective and subjective attributes, *Production Planning & Control* **9** (2), 189–200, (1998).
14. C. Parkan and M.-L. Wu, Decision-making and performance measurement models with applications to robot selection, *Computers & Industrial Engineering* **36** (3), 503–523, (1999).
15. Y. Cha and M. Jung, Satisfaction assessment of multi-objective schedules using neural fuzzy methodology, *International Journal of Production Research* **41** (8), 1831–1849, (2003).
16. T.-Ch. Chu, Facility location selection using fuzzy TOPSIS under group decisions, *International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems* **10** (6), 687–701, (2002).
17. H. Deng, C.-H. Yeh and R.J. Willis, Inter-company comparison using modified TOPSIS with objective weights, *Computers & Operations Research* **27** (10), 963–974, (2000).

18. C.-M. Feng and R.-T. Wang, Considering the financial ratios on the performance evaluation of highway bus industry, *Transport Reviews* **21** (4), 449–467, (2001).
19. J.P. Brans, P. Vincke and B. Mareschal, How to select and how to rank projects: The PROMETHEE method, *European Journal of Operational Research* **24**, 228–238, (1986).
20. B. Roy, Clessement et choix en presence de criteres multiples, *RIRO* **8**, 57–75, (1968).
21. T.L. Saaty, A scaling method for priorities in hierarchical structures, *Journal of Mathematical Psychology* **15** (3), 234–281, (1977).
22. D.L. Olson, Comparison of three multicriteria methods to predict known outcomes, *European Journal of Operational Research* **130**, 576–587, (2001).
23. W. Edwards, How to use multiattribute utility measurement for social decisionmaking, *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-7** (5), 326–340, (1977).
24. W. Edwards and F.H. Barron, SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement, *Organizational Behavior and Human Decision Processes* **60**, 306–325, (1994).
25. D.L. Olson and V.K. Dorai, Implementation of the centroid method of Solymosi and Dombi, *European Journal of Operational Research* **60** (1), 117–129, (1992).
26. P. Sprent, *Applied Nonparametric Statistical Methods*, Chapman and Hall, New York, (1989).