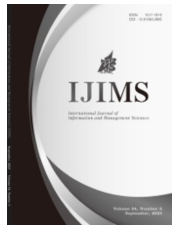
http://ijims.ms.tku.edu.tw/main.php

International Journal of
Information
and
Management
Sciences

International Journal of
Information and Management Sciences

35 (2024), 175-190. DOI:10.6186/IJIMS.202406_35(2).0005



TOPSIS Sorting with Optimized Cutoff Values

Chichang Jou[*]
Department of Information Management, Tamkang University

Keywords

Multiple criteria decision making; TOPSIS; Ordered sorting; Cutoff value.

Abstract

Many real-world decision problems require sorting alternatives into ordered classes, and often they involve multiple measures, making them multi-criteria sorting problems. Previous research on applying TOPSIS (The Technique for Order of Preference by Similarity to Ideal Solution) to these practical problems has focused on proposing criteria weights and computing relative closeness, obtained by comparing distances of alternatives to the positive and negative ideal solutions. However, the issue of how to determine the cutoff values has not been attacked before. We propose a general approach to determine optimized cutoff values, with objective weights for the TOPSIS sorting process. These cutoff values are obtained by minimizing the sum of deviations for randomly selected representative alternatives of neighboring classes. The procedure is demonstrated using two public datasets. It is then analyzed and compared with previous research and traditional data mining techniques, and the results demonstrate that TOPSIS is an effective tool for ordered sorting.

1. Introduction

The objective of many real-world sorting decisions is to divide alternatives into ordered classes. When there is only one measure involved, it is simply a matter of defining cutoff values among classes. If there are multiple measures involved, then a multiple criteria method is needed.

Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) was introduced to handle real world multi-criteria decision making (MCDM) problems (Hwang & Yoon, 1981). It helps decision maker(s) (DMs) conduct analysis, comparisons, and rankings of available alternatives when multiple criteria are involved. When it was introduced, there were not many works on sorting problems.

We first review published works involving sorting and classification work in TOPSIS. We then propose a generalized approach based on the concept of minimizing the sum of deviations for

*corresponding author. Email: cjou@mail.tku.edu.tw

randomly selected representative alternatives of neighboring classes to determine optimized cutoff values and to then exploit ranking indices, distance functions, and weights. An integrated ordered sorting procedure is presented and demonstrated using the qualitative bankruptcy dataset from UCI and the economic freedom index of countries from the Heritage Foundation.

Vincke (1992) defined a sorting problem as an operation dividing alternatives into groups of alternatives according to some norms. The problem could be further divided into classification and clustering in data analytics (Olson & Delen, 2008), depending on whether the number of groups is known *a priori*. Sorting problems can concentrate on obtaining ordered classes according to the ranking indices from the aspect of MCDM. Doumpos and Zopounidis (2002) identified many practical applications in medicine, pattern recognition, marketing, environmental management, financial management, etc. In the presence of big data, the sorting problem has clearly grown in importance. For instance, sorting analysis could give retailers unprecedented insights into customer behavior, thus allowing them to improve their customer experience, like product recommendations, personalized search, etc. Other applications include ABC analysis, medical diagnosis, and business failure prediction (Kadziński et al., 2021). An effective MCDM ordered sorting algorithm is thus needed.

Although many researchers extend TOPSIS to handle uncertainties, for example fuzzy TOPSIS (Salih et al., 2019), we focus on traditional preference-based framework of TOPSIS to emphasize on the determination of optimized cutoff values of neighboring classes. We shall demonstrate how TOPSIS, a distance-based compromise solution, can support sorting with optimized cutoff values in an objective and systematic way.

2. Literature Review

Wu and Olson (2006) were the first researchers that utilized TOPSIS for binary sorting, i.e. loan credit rating, taking a cutoff value equivalent to its relative closeness obtained from the training data. Weights on criteria were attained from data by ordinal least squares regression. In comparison with the results of decision tree algorithms, TOPSIS gave a better fit to test data from banking loan cases in Canada. Simulation was employed to examine the effects of perturbation data on the results.

Chen et al. (2007) exploited distance measures to fictitious ideal cases, similar to TOPSIS, to classify water usage of Canada's 1285 municipalities. The data of three groups of representative cities, 12 cities in total, were analyzed with quadratic programming. Their objective function minimized total error defined as deviations from group cutoff values, and the model identified cutoff values of groups and criteria weights. All these municipalities were subsequently classified into three types: robust systems, low-risk systems, or high-risk systems regarding water management.

Li et al. (2011) employed the concept of TOPSIS with case-based reasoning on known solutions, which are similarities to positive and negative ideal cases for binary business failure prediction. According to their experiments on datasets in China, the results showed that the proposed approach could generate better discriminating capability in a normal economic environment, but not in a financial crisis environment.

Zhu et al. (2013) proposed C-TOPSIS classification for credit rating of applicants. From the training data, a 2-norm Euclidean distance function derived criteria weights, yielding a cutoff value for test applications. C-TOPSIS results on accuracy, computational efficiency, and interpretability were compared with those of seven popular data mining algorithms through two credit datasets. Only the proposed method ranked among the top 3 in all of the above three aspects, thus demonstrating its advantages.

Sabokbar et al. (2016) presented a novel sorting method, TOPSIS-SORT, for evaluating Tehran environmental quality. Twenty-two districts were sorted into five groups, whose profiles and cutoff values on criteria were supplied by experts. Ouenniche et al. (2018) initiated an integrated in-sample and out-of-sample evaluation framework for bankruptcy prediction. Their key step was to find the upper and lower bounds of the TOPSIS score-based cutoff values, and they identified optimal values through non-linear search algorithms. Empirical results on a UK dataset of bankrupt and non-bankrupt firms showed outstanding prediction performance.

de Lima Silva and de Almeida Filho (2020) considered two versions of cutoff values for ordered sorting: TOPSIS-Sort-B is based on the boundary profile from Sabokbar et al. (2016), and TOPSIS-Sort-C is for the characteristic profile allowing upper and lower bounds on boundary profiles. Cutoff values on each criterion of both versions were delivered by experts to start their algorithms. After exploring the case of world economic freedom in 180 countries, the similarity percentage or correct ratio of both versions was over 89%. de Lima Silva et al. (2020) went on to propose PDTOPSIS-Sort, a preference disaggregation TOPSIS approach for sorting in which non-linear programming was used to obtain boundary profiles and weights on criteria. The objective function was to minimize the sum of the error variables of the reference alternatives in their corresponding classes, following Doumpos and Zopounidis (2002). The approach was applied to a 50-corporate bond classification problem with ten criteria given by nine reference alternatives in three classes.

Some published research studies have applied sorting to VIKOR (Opricovic, 1998), a distance-based compromise MCDM method. Baccour (2018) set up a combined TOPSIS and VIKOR method called ATOVIC classification and applied it to five UCI medical datasets. That work separated these datasets into reference and test sets and used the core concept of VIKOR for classification, i.e., the correlations among three indices.

Demir et al. (2018) offered VIKORSORT for classifying 20 green suppliers into three groups. Using pre-assigned limit profiles on criteria, they utilized weighted distance to the ideal solution for the evaluation in classification. Yamagishi and Ocampo (2022) applied TOPSIS-SORT (Sabokbar et al., 2016) to classify the degree of exposure of customers to COVID-19 in 40 restaurants in the Philippines. Two cutoff values on attributes for three classes are subjectively determined.

In summary, core works on handling multiple cutoff values could be acquired by expert judgments, reference alternatives/cases, or training data. Some authors claimed that the sorting results were successful, but their results were rather dependent on the cases. This gap in an objective and efficient method of TOPSIS sorting problems is what our research is looking to fill.

3. Proposed Approach

Inspired by the concept of minimizing total deviations, we employ Chen et al.'s concept (2007) to obtain a reliable way to determine cutoff values for ordered sorting.

3.1 Determination of optimized cutoff values

The determination of optimized cutoff values for multiple classes is a challenging task. These values are commonly obtained from expert judgments, reference alternatives/cases, or training data. The first approach, through expert judgments, is useful, but it is difficult to verify the results in different environments or by different expert teams. The second and third approaches employ a similar concept in which the cutoff values can be procured from the alternatives in the different classes.

Chen et al. (2007) utilized case-based distance models with applications in water resource management. They first transformed the multi-criteria alternatives into a consequence data space, aggregating criteria performance and their corresponding weights. They then proposed an algorithm that calculated the distance of each alternative from the fictitious ideal alternative, mimicking the separation measure to the positive ideal solution (PIS) in TOPSIS, and ranked the alternatives in ascending order. The performance of the representative alternatives with positive or negative tolerances (i.e., errors) is imbedded into the constraint set to form the boundaries of the classes. Their objective function is to minimize total squared errors by counting the deviations of the alternatives away from the boundaries. In addition, there are two types of deviations from the boundaries. For two neighboring classes, one class of the alternatives has deviations less than a fixed cutoff value and the other class greater than the fixed cutoff value. The cutoff values and the corresponding weights on criteria could be successfully acquired by solving this quadratic programming problem.

This method (Chen et al., 2007) provides an objective approach to multi-class problems. However, after examining other cases, we find Chen et al.'s technique (2007) difficult to follow. The first difficulty is in obtaining multiple cutoff values. Their algorithm frequently stops with just one cutoff value. This means that the program searches for the boundary of only two classes. The second difficulty is about the resulting criteria weights. Almost half of the obtained criteria weights are zeros. These unbalanced weights create a challenge to implement them in real-life situations.

Based on these observations, we employ the concept of minimizing the sum of squares of the residuals or deviations (Johnson & Bhattacharyya, 2019) to determine the optimized cutoff values. Our proposed program seeks a cutoff value between two neighboring classes in which the corresponding representative alternatives are randomly selected. To avert the bias of DMs' preferences, we acquire the weight sets by applying several objective weighting approaches (Zardari et al., 2015). Since we do not assume much information on the characteristics of the target dataset, the trained and test portions of the dataset could be utilized to examine which weighting approach is more suitable for the dataset. By adopting various weight sets, the results on the misclassified alternatives are used to validate the effectiveness of these weight sets. Section 3.2 gives details about weight set selection.

To explain the details of our approach, suppose the number of classes is L, and there are m alternatives and n criteria. A cutoff value R_g, $g = 1,\dots,L-1$, is to be determined for the g^{th} and the $(g+1)^{th}$ classes. We randomly select t_g numbers of alternatives for the g^{th} class and t_{g+1} numbers of alternatives from the $(g+1)^{th}$ class. Let α_g^i, $i=1,\dots,t_g$, denote an upper-bound error for the i^{th} selected alternative in the former class, where $-1 \leq \alpha_g^i \leq 0$; and β_{g+1}^i, $i=1,\dots,t_{g+1}$, denote a lower-bound error for the i^{th} selected alternative in the latter class, where $0 \leq \beta_{g+1}^i \leq 1$. Let r_{ij}, $i=1,\dots,m$ and $j=1,\dots,n$, be the normalized performance values from the dataset; $r_j^+ = max_i(r_{ij})$ for the benefit criteria; and $r_j^+ = min_i(r_{ij})$ for the cost criteria. Suppose the weight for criterion j is w_j. The weighted normalized performance measure of alternative i from the ideal solution can be represented as $\sum_j w_j |r_{ij} - r_j^+|$. The suggested programming for two neighboring classes, g and $g+1$, is:

$$\text{Minimize d} = \sum_{i=1}^{t_g} (\alpha_g^i)^2 + \sum_{i=1}^{t_{g+1}} (\beta_{g+1}^i)^2 \quad (3.1)$$

Subject to

$$\left(\sum_j w_j |r_{ij} - r_j^+|\right) + \alpha_g^i \leq R_g,\ g=1,\dots,L-1 \quad (3.2)$$

$$\left(\sum_j w_j |r_{ij} - r_j^+|\right) + \beta_{g+1}^i \geq R_g,\ g=1,\dots,L-1 \quad (3.3)$$

$$-1 \leq \alpha_g^i \leq 0,\ g=1,\dots,L-1 \quad (3.4)$$

$$0 \leq \beta_{g+1}^i \leq 1,\ g=1,\dots,L-1 \quad (3.5)$$

$$R_g \leq 1,\ g=1,\dots,L-1 \quad (3.6)$$

$$\sum_j w_j = 1, \quad (3.7)$$

$$w_j \geq 0. \quad (3.8)$$

After assigning a weight set to the above program, the cutoff value R_g of classes g and $g+1$ is obtained. If there are more than two classes, then starting from R_1, with randomly selected representative alternatives from the pair of next neighboring classes, the process continues until all neighboring classes are examined.

Figure 1 uses two criteria X_1 and X_2 to illustrate the above optimized cutoff value determination procedure. In Fig 1, v_1^+/v_1^- is the largest/smallest X_1 value, and v_2^+/v_2^- is the largest/smallest X_2 value. The ellipse forms the set of possible alternatives. Moreover, A^+ is the PIS and A^- the NIS. The three alternatives above the cutoff value R_g are the representative cases of class g, while the three below are the representative cases of class $g+1$. Through randomly selected representative alternatives on both sides, the program searches for an optimized cutoff value R_g satisfying goal (3.1) subject to constraints (3.2) to (3.8). The ranking index of TOPSIS is relative closeness C_i^*, for $i = 1, \dots, m$ alternatives and $0 \leq C_i^* \leq 1$, and we mark its middle value $C_i^*=0.5$ with a red solid line as a reference. The cutoff value R_g acts in the same role as C_i^* for TOPSIS decision, where $0 \leq R_{g+2} \leq R_{g+1} \leq R_g \leq 1$ in Fig. 1. The line R_g is solid purple, and lines for R_{g+1} and R_{g+2} are dotted purple. In addition, TOPSIS evaluation relies on the aggregation of preference values under multiple criteria and is a holistic approach, which is different from the disaggregation approach of Kadziński et al. (2021).

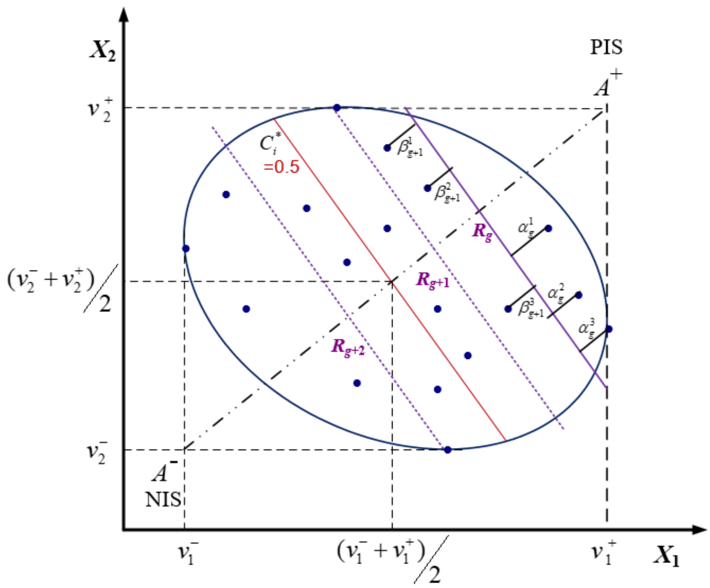


Figure 1 *Optimal Cutoff Value Determination*

3.2 Weight set selection

Another important issue in MCDM is to determine the weights of criteria. For any selected dataset, we might not have enough information on relative preference for criteria. In such a case, subjective weighting beyond applying equal weights could be an arduous task. As a consequence, objective weights generated from the data are useful to capture the characteristics of the data. This study adopts three objective weights: the entropy method, the standard deviation method, and the CRiteria Importance Through Intercriteria Correlation (CRITIC) method (Zardari et al., 2015). Together with equal weights, for cases when there is no preference on the criteria, we will compare the error rates of the sorting problem under these four weight sets to pick one with the least error rate. In such a design, TOPSIS sorting will be managed systematically.

3.3 Proposed sorting procedure

In the original TOPSIS, the ranking index for alternative i is its relative closeness C_i^*, which is a combination of two separation measures S_i^+ and S_i^-, using an n-dimensional Euclidean distance. Since we do not know the characteristics of the targeted dataset in advance, some variants in the TOPSIS algorithm could be considered in verification so as to look for a better result. Note that S_i^+ and S_i^- are opposite concepts in assessment. Based on the concept of Wu and Olson's data mining method (2006), we propose a more general sorting procedure in Figure 2. The steps in Figure 2 are explained as follows.

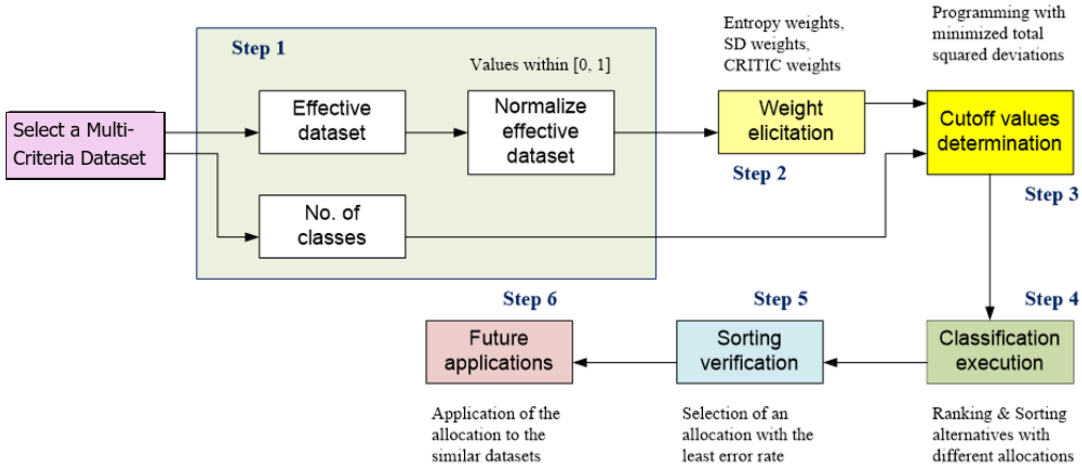


Figure 2 *Flowchart of the Suggested Sorting Procedure*

Step 1: Gather data.

1.1) Select a classification decision matrix or dataset.

Depending upon the characteristics of the data, organized or unorganized, if the data are unorganized, then the data will be statistically tested for significance. The selected decision matrix, with m alternatives (or instances) and n criteria (or columns), will be linearly normalized, which means the data are divided by their maximal value for benefit criteria and their minimal value is divided by the data for cost criteria, as follows:

$$\boldsymbol{R} = \begin{array}{c c} & \begin{array}{cccccc} X_1 & X_2 & \cdots & X_j & \cdots & X_n \end{array} \\ \begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_i \\ \vdots \\ A_m \end{array} & \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1j} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2j} & \cdots & r_{2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ r_{i1} & r_{i2} & \cdots & r_{ij} & \cdots & r_{in} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mj} & \cdots & r_{mn} \end{bmatrix}_{m \times n} \end{array} \quad (3.9)$$

where i=1,…,m and j = 1,…,n.

1.2) Obtain the number of classes. If this number is unknown, then try the same procedure using different number of classes.

Step 2: Weight elicitation.

Based on the contents of the normalized matrix or dataset of Eq. (3.9), many candidate weight sets could be acquired. As mentioned in Section 3.2, we consider the entropy method, the standard deviation method, the CRITIC method, and equal weights.

Step 3: Cutoff values determination.

The program, equations (3.1) through (3.8), is executed to determine the cutoff values for multiple classes. If there are L classes to be segregated, then L-1 cutoff values need to be determined. For each weight set, with the randomly selected representative alternatives in two neighboring classes, we run the program L-1 times.

Step 4: Classification execution.

4.1) Rank alternatives by TOPSIS.

According to the TOPSIS ranking index, including relative closeness C_i^* or separation measures S_i^+ and S_i^-, of alternatives i, $i = 1,\ldots,m$, allocate all alternatives in descending order, respectively. Note that:

$$C_i^* = \frac{S_i^-}{S_i^+ + S_i^-} \tag{3.10}$$

where S_i^+ is the distance between alternative i and PIS, and S_i^- is the distance between alternative i and NIS. Besides Euclidean distance, Manhattan distance is also considered as an option for the analysis.

4.2) Sort the alternatives into classes.

Consider the obtained cutoff values as the boundaries of the classes and compare the ranking indices of TOPSIS for the classification. Allocate all alternatives to the appropriate ordered classes. There are four groups of allocations from the four weight sets.

Step 5: Sorting verification.

Count the number of misplaced alternatives in each class, and enumerate the error rate of each allocation. Compare four groups of allocations, and select one with the lowest error rate.

Step 6: Future applications.

Since the choice is the classification with the least error rate, the corresponding allocations (cutoff values and weight set) could be used for classifying datasets with similar characteristics in the future.

4. Analysis

To demonstrate the effectiveness of our approach, we present its application on two public datasets. The first one gives categorical data on bankruptcy, taken from the UCI Machine Learning Repository at https://archive.ics.uci.edu/ml/datasets/qualitative_bankruptcy. This application needs only one cutoff value for two classes, yes or no. The second dataset deals with the Index of Economic Freedom data from the Heritage Foundation at https://www.heritage.org/index/. This application needs four cutoff values for categorizing countries in the world into five classes.

4.1 Corporate bankrupt data

Bankruptcy is a legal process by which individuals or other entities that are unable to repay their creditors seek relief from some or all of their debts. Corporate bankruptcy is a serious situation for managing businesses and for which stakeholders would like to predict. In this data set, there are 250 instances with seven attributes, where the last one is the dependent variable: Industrial Risk (P, A, N), Management Risk (P, A, N), Financial Flexibility (P, A, N), Credibility (P, A, N), Competitiveness (P, A, N), Operating Risk (P, A, N), and Class (B, NB). The letters in the parentheses under the attributes are presented as: P = Positive, A = Average, N = Negative, B = Bankruptcy, and NB = Non-Bankruptcy.

The categorized data are linearized to obtain cardinal data. To avoid extreme values of the ordinal representation on independent attributes, we set P = 0.75, A = 0.5, and N = 0.25. The values of the dependent attribute, the two outcome classes, are set as NB = 1 and B = 0. Table 1 lists the weights of the six criteria of the four weight sets from the dataset.

Table 1 *Four Weight Sets for All Criteria*

Weight Set	Criteria					
	Industrial Risk	Management Risk	Financial Flexibility	Credibility	Competitiveness	Operating Risk
Entropy Weights	0.1496	0.1733	0.1688	0.1554	0.1728	0.1802
Standard Deviation Weights	0.1636	0.1633	0.1597	0.1653	0.1752	0.1728
CRITIC Weights	0.1934	0.1699	0.1474	0.1497	0.1423	0.1973
Equal Weights	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667

The dataset has ratios of 70% NB and 30% B. We use three levels of sampling respectively; i.e., 2%, 6%, and 10% of the total instances. Cutoff values, obtained using equations (3.1) through (3.8), are averaged from three sampling results. Whenever the decision makers are afraid of an unstable result obtained through random sampling, multiple samplings could be applied. Fig. 3 shows that for the three levels of sampling, the resulting error rate of sorting under each weight set. Their cutoff values range from 0.4833 to 0.4871.

We could see from Figure 3 the sorting performances are rather good. The error rates for cases with a 2% sample size (using only 5 of 250 data points) are a little worse than those of 6% and 10% sample sizes. Among the weight sets, the error rates by CRITIC weights have inferior results. Because CRITIC weights consider the correction among the attributes, this means that there could be less correction among the data.

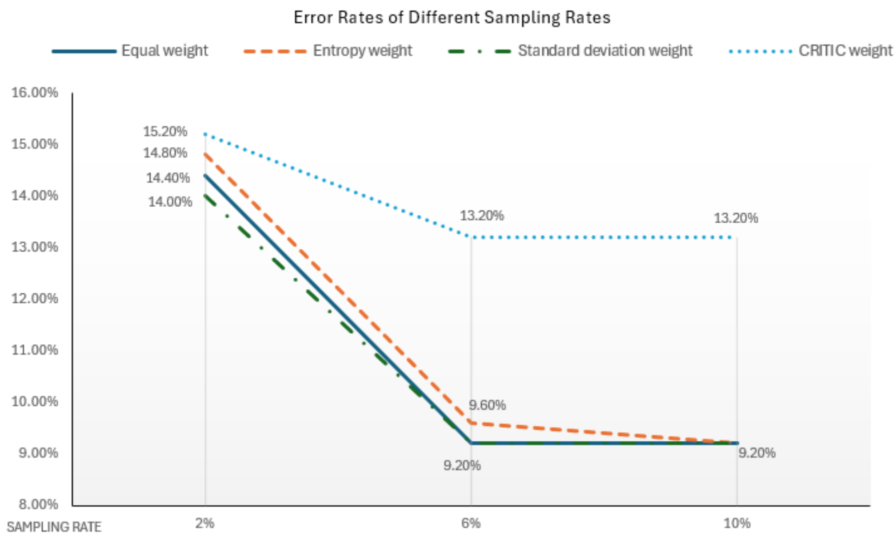


Figure 3 *Error Rates of Four Weights Sets with Different Sampling Rates*

On the same dataset, Aruldoss et al. (2015) applied the ant-miner algorithm and obtained 96.2 correc t rate; while Koklu and Tutuncu(2014) applied Naive Bayes Classifier, Multilayer Perceptron, J48 and Classification via Regression, and obtained correct rates of 96.57%, 94.86%, 95.43% and 96.00% separately. We also utilize the sklearn decision tree classifier in https://scikit-learn.org on the same dataset. Setting max_depth = 5 and default values for the remaining parameters, the correct rate reaches 96.00%. Th above results are a little better than our best results of 90.80%. This analogy demonstrates that TOPSIS does indeed have good performance for sorting.

4.2 Index of Economic Freedom data

To further demonstrate the effectiveness of our approach, we present data from The Heritage Foundation's Index of Economic Freedom, using the 2021 report. This index measures political bodies in terms of trade freedom, tax burden, judicial effectiveness, and other metrics. It is designed to be a composite measure of the quality of political economic institutions. Our intent is not to challenge or dispute The Heritage Foundation system, but rather to numerically demonstrate how TOPSIS sorting works.

Table 2 (Step 2) lists the weights of the 12 criteria from the four weight sets.

Using the data from year 2021, the four sets of cutoff values, from the averaged values of six randomized selected sets of alternatives, are illustrated in Table 3.

Table 2 *Weights on Criteria from Different Weight Sets (Year 2021)*

Weight Set	Entropy Weights	Std. Dev. Weights	CRITIC Weights	Equal Weights
Property Rights	0.0881	0.0891	0.0686	0.0833
Judicial Effectiveness	0.1243	0.0954	0.0791	0.0833
Government Integrity	0.1286	0.1004	0.0850	0.0833
Tax Burden	0.0247	0.0560	0.0796	0.0833
Government Spending	0.1043	0.0985	0.1618	0.0833
Fiscal Health	0.1487	0.1195	0.1425	0.0833
Business Freedom	0.0472	0.0741	0.0591	0.0833
Labor Freedom	0.0429	0.0692	0.0698	0.0833
Monetary Freedom	0.0309	0.0604	0.0574	0.0833
Trade Freedom	0.0268	0.0562	0.0462	0.0833
Investment Freedom	0.1192	0.0937	0.0799	0.0833
Financial Freedom	0.1143	0.0875	0.0710	0.0833

Table 3 *Cutoff Values and Error Rates for Four Weight Sets by Euclidean Distance (Year 2021)*

Weight Set	Cutoff Values				Error Rate		
	R_1	R_2	R_3	R_4	C_i^*	PIS(S_i^+)	NIS(S_i^-)
Entropy Weights	0.81790	0.71338	0.58320	0.48975	27.53%	58.43%	64.04%
Standard Deviation Weights	0.82276	0.72690	0.61195	0.52130	22.47%	62.36%	62.92%
CRITIC Weights	0.80747	0.72981	0.63390	0.55585	46.63%	46.63%	80.34%
Equal Weights	0.82439	0.73499	0.62891	0.53803	12.92%	56.18%	69.10%

After obtaining cutoff values, we employ them for sorting countries (Step 4). In compliance with the ranking indices, relative closeness C_i^*, in the original TOPSIS, we also make use of separation measures for PIS (S_i^+) and NIS (S_i^-) in the evaluation. Their four error rates, counted by C_i^*, appear in the 6th column of Table 3. Here, equal weights provide the best result (12.92%). The corresponding error rates for PIS(S_i^+) and NIS(S_i^-) are also listed in the 7th and 8th column for comparison. The ranking index obtained from the two separation measures does not yield good results. Compared to allocation by equal weights, entropy weights and standard deviation weights provide lower error rates (Step 5), and the CRITIC weight method yields the greatest error rates.

To evaluate the effectiveness of our procedure, we employ these weights and cutoff values to classify the data of 2020 economic freedom (Step 6). The corresponding error rates are in Table 4 and appear slightly worse than the results for the year 2021. The former results show that entropy weights have a lower error rate, 29.44%, but are higher than that of equal weights, 16.67%. Note that our error rate is close to the results of de Lima Silva and de Almeida Filho (2020), which are 10.56%, 22.22%, and 10.00% for directly using year 2020 data.

Table 4 *Error Rates from Different Weight Sets by Euclidean Distance (Year 2020)*

Weight Set	Error Rate			Note
	C_i^*	PIS(S_i^+)	NIS(S_i^-)	
Entropy Weights	29.44%	68.89%	64.04%	Four cutoff values are shown in Table 7
Standard Deviation Weights	30.00%	69.44%	62.92%	
CRITIC Weights	45.56%	55.56%	61.67%	
Equal Weights	16.67%	63.33%	69.10%	Reference

Another variant is employing Manhattan distance for analysis. The corresponding results are in Table 5 for the year 2021 data. Standard deviation weights yield a better outcome with an error rate of 11.24%, which is still inferior to 9.55% from equal weights. The distance measure to PIS also produces rather good results. Table 6 displays the errors after applying the same cutoff values to the data for the year 2020. In general, standard deviation weights deliver better results with a 13.33% error rate, which is inferior to 7.22% from equal weights. Here, the error rates by the distance to PIS generate a superior outcome at 3.33%, much better than de Lima Silva and de Almeida Filho's result (2020). These results verify that TOPSIS with Manhattan distance and SAW (Simple Additive Weighting) share similar characteristics (Hwang & Yoon, 1981).

Table 5 *Error Rates from Different Weight Sets by Manhattan Distance (Year 2021)*

Weight Set	Error Rate			Note
	C_i^*	PIS(S_i^+)	NIS(S_i^-)	
Entropy Weights	21.91%	16.85%	26.40%	Four Cutoff Values are Shown in Table 7.
Standard Deviation Weights	11.24%	9.55%	25.28%	
CRITIC Weights	22.47%	27.53%	25.28%	
Equal Weights	9.55%	10.11%	21.35%	Reference

We also utilize the sklearn decision tree classifier on the 2021 dataset. Setting max_depth = 5 and default values for the remaining parameters, the correct rate is 66.67%, which is worse than our best results of 88.76% for the standard deviation weights with Manhattan distance. This analogy demonstrates that TOPSIS performs better than decision tree classifier for multi-class sorting. We think the difference is because decision tree treats the soring results in multi-class sorting as categorical classes, while TOPSIS captures more nature of distance in multi-class ordered sorting.

Table 6 *Error Rates from Different Weight Sets by Manhattan Distance (Year 2020)*

Weight Set	Error Rate			Note
	C_i^*	PIS(S_i^+)	NIS(S_i^-)	
Entropy Weights	20.56%	15.00%	32.22%	Four Cutoff Values are Shown in Table 7
Standard Deviation Weights	13.33%	7.78%	25.00%	
CRITIC Weights	22.22%	23.89%	31.11%	
Equal Weights	7.22%	3.33%	23.33%	Reference

5. Discussion

As noted in Section 2, the difficulty for TOPSIS in ordered sorting problems is in determining the cutoff values. The experts from the Heritage Foundations selected the four cutoff values of 0.8, 0.7, 0.6, and 0.5 for their SAW method. However, TOPSIS employs the value of relative closeness for ranking. Hence, in Table 7, these four values for SAW are transformed into the corresponding cutoff values, which vary depending on the given weight sets, with their error rates for the 2021 and 2020 datasets. We observe that these error rates are much better than our proposal, except for the CRITIC weights. Equal weights produce the highest performance in terms of error rate at 6.74% and 6.11%, respectively, which is much better than the result of de Lima Silva and de Almeida Filho (2020). This evidence shows that if we can precisely determine the cutoff values, then TOPSIS indeed is an effective sorting tool.

Table 7 *Cutoff Values and Error Rates from Different Weight Sets by the Original TOPSIS*

Weight Set	Cutoff values				Error rate		Note
	R_1	R_2	R_3	R_4	2021	2020	
Entropy Weights	0.83526	0.69411	0.57285	0.46739	19.66%	21.11%	
Standard Deviation Weights	0.81887	0.70370	0.59587	0.49595	11.24%	12.22%	
CRITIC Weights	0.80096	0.67058	0.60271	0.51714	60.11%	55.00%	
Equal Weights	0.81826	0.71280	0.62066	0.52462	6.74%	6.11%	Reference

The traditional TOPSIS algorithm employs vector normalization for performance and Euclidean distance for comparing the distance of the alternatives to their PIS/NIS. For a large size of instances, vector normalization could be a burden for applications. Thus, many works make use of linear normalization, as Wu and Olson (2006) did, to have an efficient calculation. The utilization of Manhattan distance instead of Euclidean distance provides the same edge.

The TOPSIS algorithm needs a weight set as an input to the process. The weight set could be assigned in a subjective or objective way. If we know the characteristics of the dataset in advance, then we can consider the weights in a subjective way; otherwise, an objective weight set might be the choice. For comparison purposes, we assume that there is no information on the target dataset and employ four common objective weighting methods. The tested result with the least error rate from one weight set will be the selection for future applications to similar datasets.

In the dataset of the economic freedom index, however, equal weights appear to do the best job. Because the original rankings are based on SAW with equal weights, the resembling procedure results in a better fit. A similar scenario occurs when the result is calculated using Manhattan distance rather than Euclidean distance.

6. Conclusions

This research proposes a procedure for TOPSIS to solve sorting problems. For problems with unknown number of classes, the procedure could be applied using different number of classes to select best results. The major difference between TOPSIS, or any MCDM method, and data mining techniques is that the alternatives should be ranked first before classification or clustering (Nemery, 2008). Thus, an ordered group is expected to be ranked first. This characteristic provides a unique feature for demonstrating ordered classes based on some ranking indices. Moreover, searching for cutoff values among classes is difficult when applying TOPSIS to sorting problems, of which most are assigned by subjective values. We use the concept of the minimized square of deviations with alternative data to find adequate cutoff values.

For future applications on a similar dataset, error rates are employed to evaluate which weight set allocation under the variants of distance functions and ranking indices is suitable. In the economic freedom example, standard deviation weights with Manhattan distance could generally supply a better result. However, equal weights generate the best match to the original work. From these results, we know that the TOPSIS concept could be enhanced to provide a satisfactory result for sorting problems under the condition that the datasets are organized by the characteristics of the problems. For future work, this framework should be performed on larger datasets. Other objective weights (Zardari et al., 2015) could be considered in this framework.

References

Aruldoss, M., Travis, M. L., & Venkatesan, V. P. (2015). A reference model for business intelligence to predict bankruptcy. *Journal of Enterprise Information Management*, 28, 186-217.

Baccour, L. (2018). Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets. *Expert Systems with Applications*, *99*, 115-125. https://doi.org/10.1016/j.eswa.2018.01.025

Chen, Y., Hipel, K. W., & Kilgour, D. M. (2007). Multiple-Criteria sorting using case-based distance models with an application in water resources management. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 37(5)*, 680-691. https://doi.org/ 10.1109/TSMCA.2007.902629

de Lima Silva, D. F. & de Almeida Filho, A. T. (2020). Sorting with TOPSIS through boundary and characteristic profiles. *Computers and Industrial Engineering, 141,* 106328. https://doi.org/10.1016/j.cie.2020.106328

de Lima Silva, D. F., Ferreira, L., & de Almeida-Filho, A. T. (2020). A new preference disaggregation TOPSIS approach applied to sort corporate bonds based on financial statements and expert's assessment. *Expert Systems with Applications, 152,* 113369. https://doi.org/10.1016/j.eswa.2020.113369

Demir, L., Akpınar, M. E., Araz, C., & Ilgın, M. A. (2018). A green supplier evaluation system based on a new multi-criteria sorting method: VIKORSORT. *Expert Systems with Applications, 114*, 479-487. https://doi.org/10.1016/j.eswa.2018.07.071

Doumpos, M. & Zopounidis, C. (2002). *Multicriteria decision aid classification methods* (Vol. 73). Springer Science & Business Media. https://doi.org/10.1007/b101986

Hwang, C. L. & Yoon, K. (1981). Multiple attribute decision making: a state-of-the-art survey. *Lecture notes in economics and mathematical systems*, *186*(1).

Johnson, R. A. & Bhattacharyya, G. K. (2019). *Statistics: principles and methods*. John Wiley & Sons.

Kadziński, M., Martyn, K., Cinelli, M., Słowiński, R., Corrente, S., & Greco, S. (2021). Preference disaggregation method for value-based multi-decision sorting problems with a real-world application in nanotechnology. *Knowledge-Based Systems*, *218*, 106879. https://doi.org/10.1016/j.knosys.2021.106879

Koklu, M., & Tutuncu, K. (2014, July). Qualitative bankruptcy prediction rules using artificial intelligence techniques. *In Proceedings of the International Conference on challenges in IT, Engineering and Technology (ICCIET'2014), Phuket, Thailand* (pp. 21-25).

Li, H., Adeli, H., Sun, J., & Han, J.G. (2011). Hybridizing principles of TOPSIS with case-based reasoning for business failure prediction. *Computers and Operations Research*, *38(2)*, 409-419. https://doi.org/10.1016/j.cor.2010.06.008

Nemery, P. (2008). An outranking-based sorting method for partially ordered categories. In *DIMACS, Workshop and Meeting of the COST Action ICO602, Université Paris Dauphine, Paris* (pp. 28-31).

Olson, D. L. & Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media. https://doi.org/10.1007/978-3-540-76917-0

Opricovic, S. (1998). Multicriteria optimization in civil engineering systems. *Faculty of Civil Engineering, Belgrade, 2(1),* 5-21.

Ouenniche, J., Pérez-Gladish, B., & Bouslah, K. (2018). An out-of-sample framework for TOPSIS-based classifiers with application in bankruptcy prediction. *Technological Forecasting and Social Change*, *131*, 111-116. https://doi.org/10.1016/j.techfore.2017.05.034

Sabokbar, H. F., Hosseini, A., Banaitis, A., & Banaitiene, N. (2016). A novel sorting method TOPSIS-SORT: an application for Tehran environmental quality evaluation. *Economics and Management*, *19(2)*, 87-104.

Salih, M. M., Zaidan, B. B., Zaidan, A. A., & Ahmed, M. A. (2019). Survey on fuzzy TOPSIS state-of-the-art between 2007 and 2017. *Computers & Operations Research, 104*, 207-227.

Vincke, P. (1992). *Multicriteria decision-aid*. John Wiley & Sons.

Wu, D. & Olson, D. L. (2006). A TOPSIS data mining demonstration and application to credit scoring. *International Journal of Data Warehousing and Mining*, *2(3)*, 16-26. http://doi.org/ 10.4018/jdwm.2006070102

Yamagishi, K. & Ocampo, L. (2022). Utilizing TOPSIS-Sort for sorting tourist sites for perceived COVID-19 exposure. *Current Issues in Tourism, 25(2)*, 168-178. https://doi.org/10.1080/13683500.2021.1918070

Zardari, N. H., Ahmed, K., Shirazi, S. M., & Yusop, Z. B. (2015). *Weighting methods and their effects on multi-criteria decision making model outcomes in water resources management*. Springer.

Zhu, X., Li, J., Wu, D., Wang, H., & Liang, C. (2013). Balancing accuracy, complexity and interpretability in consumer credit decision making: A C-TOPSIS classification approach. *Knowledge-Based Systems*, *52*, 258-267. https://doi.org/10.1016/j.knosys.2013.08.00

Chichang Jou
Associate Professor, Department of Information Management, Tamkang University, Taiwan, ROC.
E-mail address: cjou@mail.tku.edu.tw
Major area(s): Soft Computing, Data Mining, Artificial Intelligence, Algorithms

(Received May 2023; accepted March 2024)

Guide to Authors

1. Submission of Manuscripts
   - All manuscripts should be submitted (as either PDF, LaTeX or Word files) through the online submission system at https://www.ipress.tw/J0177.
   - Three files are necessary. The first is a regular manuscript. The second keeps the same as the first except all informations related to author(s) are removed, including acknowledgment. The third is Authors' Information (see item 12).
   - Papers submitted to this journal must NOT been previously published nor be under review by another journal. Any form of duplicate publication or plagiarism is absolutely prohibited. Violation of it will be penalized. Author(s) alone should take all responsibilities of all possible accusations due to violation of those previously mentioned.
   - Any addition, deletion, or rearrangement of the author's name in the author list should be done before the manuscript is accepted and approved by the journal editor. Please consider the list and order of authors carefully before submitting.
   - Once the files are deemed to have complied with the **APA Formatting and Citation (7th Ed.)** requirements, an acknowledgement email will be send together with a paper ID to the corresponding author.
2. Form of Manuscripts
   - The length limit for each manuscript is 20 pages including figures, tables and others.
   - Manuscript should use 12-point Times New Roman font and 1.5 spacing throughout. Space should be fully utilized.
   - All illustrations, photographs, tables, etc., should be on separate sheets, and should be included in each copy. Each page of the manuscript should be numbered.
   - Manuscript should be written in impeccable English (either US or UK spelling is accepted, but not a mixture of both).
3. Organization of the Paper
   - HEADING: The title of the paper should be concise and informative. Successive lines should give the author's name, academic or professional affiliation, and address.
   - ABSTRACT: Every manuscript must start with a concise abstract, no more than 150 words, followed by no more than 5 keywords.
   - MAIN TEXT: Structured clearly with sections such as (but not limited to) Introduction, Literature Reviews, Notations and Assumptions, Data and Methodology, Model Formulation, Research Design, Data Analysis and Results, Discussion, Conclusions, etc.
   - ACKNOWLEDGEMENTS: This section includes acknowledgements of assistances and financial support, etc. Please note that only accepted manuscripts may include this section, else place this section in your authors' file.
   - REFERENCES: See detailed instructions in item 11.

4. Illustrations
    - All Illustrations should be submitted in a form suitable for reproduction. Number the illustrations according to the sequence of their appearance in the text, where they are to be referred to as "Fig. 1", "Fig. 2", etc. Each illustration may have a legend, if required.
    - Photographs should be glossy prints. The authors name and figure number should be indicated on the back of each illustration.
5. In-text Citations
    - In the text, follow the rules implied in these examples. For one authr use: (Foxall, 2018) Or According to Foxall (2018).
    - For two authors use: (Mason & Missingham, 2019) Or According to Mason and Missingham (2019).
    - For more than three authors use: (Ewert et al., 2014) or According to Ewert et al. (2014).
6. Tables
    - Tables should be typed on separate pages. Number the illustrations according to the sequence of their appearance in the text, where they are to be labeled as "Table 1", "Table 2", etc.
    - Table titles should be short and self-explanatory.
    - A brief title should be given above each table, and any footnotes below (see Section 8).
7. Headings
    - The following sequence of headings should be used: 1, 1.1 (then 1.2, 1.3, ...), 1.1.1 (then 1.2, 1.1.3, ⋯).
    - Please refrain from using fourth level sections/headers.
8. Footnotes
    - In the text, footnotes are not permitted. They should be properly included in the context.
    - Use smaller font size for table footnotes.
9. Symbols and Abbreviations
    - Please use widely accepted symbols and forms of abbreviation.
    - If there is any doubt in your mind about a particular symbol or abbreviation, give the full expression followed by the abbreviation, when it appears in the text for the first time.
10. Mathematics
    - Mathematical expressions and equations should be properly typewritten, with all symbols aligned as they are to appear in print.
    - All Greek letters and other special symbols must be identified.
    - Vectors will be set in bold face and should be indicated in the manuscript by underlining with a wavy line.

- Equations or formulae should be numbered serially on the right-hand side by Arabic numerals in parentheses. For example, the first formula in Section 3 is numbered by (3.1). Only equations explicitly referred to in the text should be numbered.

11. References

- All literature citations should be collected in a list at the end of the paper and numbered alphabetically according to the last name of the first author. Retain the original title for publications in languages using the Roman alphabet. However, those employing Cyrillic and other non-Roman alphabets should be transliterated. Please note the original language at the end, e.g., "(in Russian)."
- Include DOI of the references whenever available.
- Please follow the style below:

(1) References to Book & eBook: Ewert, E. W., Mitten, D. S., & Overholt, J. R. (2014). *Natural environments and human health*. CABInternational. https://doi.org/10.1079/978184593 9199.0000 (Note: Include the DOI using the format https://doi.org/10.xxxx/xxxx)

(2) References to Chapter in an edited book or eBook: Aron, L., Botella, M., & Lubart, T. (2019). Culinary arts: Talent and their development. In R. F. Subotnik, P. Olszewski-Kubilius, & F. C. Worrell (Eds.), *The psychology of high performance: Developing human potential into domain-specific talent* (pp. 345-359). American Psychological Association. https://doi.org/10.1037/0000120-016 (Note: Use this format for both print and eBook edited book chapters, including edited book chapters from academic research databases. Include DOI if online. Please also take note how the names are given for the Editors, the initials come first.)

(3) References to Article from research databases: Washington, E. T. (2014). An overview of cyberbully in higher education. *Adult Learning*, 26(1), 21-27. https://doi.org/10.1177/1045159514558412 (Note: For journals, we italicise the Journal Name and Volume, instead of the article title. Always find articles from prominent research databases instead of some random websites, as they may be predatory or fake.)

(4) References to Conference proceedings published in a journal: Duckworth, A. L., Quirk, A., Gallop, R., Hoyle, R. H., Kelly, D. R., & Matthews, M. D. (2019). Cognitive and noncognitive predictors of success. *Proceedings of the National Academy of Sciences*, USA, 116(47), 23499-23504. https://doi.org/10.1073/pnas.1910510116 (Note: Normally Journals will indicate the Proceedings title, if not given, then please cite the source like a Journal with DOI as shown above instead of conference proceedings.)

(5) References to Report by a Government Agency: National Cancer Institute. (2019). Taking time: *Support for people with cancer* (NIH Publication No. 18-2059). U.S. Department of Health and Human Services, National Institutes of Health. https://www.cancer.gov/publications/patient-education/takingtime.pdf(Note: The specific agency responsible for the report appears as the author. The link to the report should be given. If publication number is given, then include it as well.)

(6) References to Unpublished Dissertation or Thesis: Harris, L. (2014). *Instructional leadership perceptions and practices of elementary school leaders* [Unpublished doctoral dissertation]. University of Virginia. (Notes: 1. A dissertation or thesis is considered as unpublished when you gain access through direct reading at specific library/repository but not accessibly publicly. If you find it online following published dissertation or thesis reference shown above. 2. When a dissertation or thesis is unpublished, include the description "[Unpublished doctoral dissertation]" or "[Unpublished master's thesis]" in square brackets after the dissertation or thesis title.)

(7) References to Webpage: Bologna, C. (2019, October 31). *Why some people with anxiety love watching horror movies.* HuffPost. https://www.huffpost.com/entry/anxiety-love-watching-horrormovies l_5d277587e4b02a5a5d57b59e (Notes: 1. You are advised to only use webpages with clear indication of authors and date of publication. 2. The name of the website (e.g. HuffPost) should be included before the URL. 3. No retrieval date is needed when date of publication is given.)

(8) You may refer to the full manual or follow the key information from https://apastyle.apa.org/style-grammar-guidelines/references/examples#textual-works.

12. Authors' Information

- Authors should provide the following informations in a separate file:

(1) Grant-supported research like funding, fellowship or any financial backing from person, university, organization or government.

(2) Authors' full affiliations, (department, university, country) email address and major area (at most 3 items) are to be given in order of authors. For example, Department of xxx, xxx University, ..., Canada.
E-mail: xxx
Major area(s): xxx
Department of xxx, xxx University, ..., USA.
E-mail: xxx Major area(s): xxx etc..