Correlated Criteria in Decision Models: Recurrent Application of TOPSIS Method

Cosmina Ioana BONDOR* and Adriana MUREŞAN

“Iuliu Haţieganu” University of Medicine and Pharmacy Cluj-Napoca, 8th Victor Babeş, 400012, Cluj-Napoca, Romania.
E-mails: cbondor@umfcluj.ro; amuresan@umfcluj.ro

* Author to whom correspondence should be addressed; Tel.: +40264 431697; Fax: +40264 593847

Received: 1 February 2012 / Accepted: 29 February 2012 / Published online: 10 March 2012

Abstract
The purpose of multicriteria decision models is to help decision maker to evaluate each alternative and to rank them in descending order of performance. A problem can appear when the criteria are not independent. This study explores the effect of multicollinearity between criteria in decision making with the technique for order preference by similarity to ideal solution (TOPSIS) and proposes an algorithm to resolve this problem. The algorithm was based on the application of the TOPSIS method several times until all the components are uncorrelated. The algorithm was applied on two examples from medical field to demonstrate its effectiveness. After we applied the purposed algorithm on two examples the index result from TOPSIS was equal correlated with all the criteria.

Keywords: TOPSIS method; Diabetes mellitus type 2; Multicollinearity.

Introduction
Problems of decision making where the evaluation of possible alternatives with different criteria are widespread today not only in the economy. To resolve a situation and choose that alternative which is optimal in terms of evaluation criteria is a problem not only important but also difficult.

We consider the multiatribut decision models because it may have applications in the medical field [1-6]. Multi-criteria methods were used in the medical fields in combination with Shannon entropy [7-8], fuzzy theory [9], artificial intelligence algorithms [10] or neural networks [11].

From the multiatribut decision making methods we chose TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method because it can be use not only to determin the best alternative, but it can be use for ranking the alternatives.

In TOPSIS method the best alternative is considered the alternative that minimizes the distance to the ideal solution. The ideal solution is the solution that maximizes all the maximum criteria and minimizes all the minimum criteria. The optimal alternative is the alternative for which the distance to ideal solution is minimal [12]. Hwang and Yoon introduced this method in 1981 [13]. TOPSIS method was used in medical field by several authors [14-20]. We found only two studies in medical field were the authors applied TOPSIS method for rank risk factors: in bronchial asthma [19] and in post kidney transplant diabetes mellitus [20]. TOPSIS method takes in account the characteristics of the factors: risk or protection factor and amplitude of their influence [20].

The assumptions of the TOPSIS method are independence and non-correlation between attributes [21]. This assumption frequently is violated and then the evaluations done using TOPSIS
in presence of multicolinearity can be wrong [22].

The aim of our study was to explore the effect of multicollinearity between two or more criteria in a decision making model with TOPSIS and to found a solution for the multicollinearity problem.

Material and Method

**TOPSIS Method**

We have, generally, m indicators (symptoms, characteristics, criteria) \( C_j, j=1,m \) of the same condition (disease, problem, state) and n alternative solutions \( V_i, i=1,n \). In the following, we present the algorithm steps [12].

- **Step 1.** Construct the matrix of consequences \( A = [a_{ij}], i=1,n, j=1,m \)
- **Step 2.** Construct the matrix of normalized consequences \( R = [r_{ij}], i=1,n, j=1,m \).
- **Step 3.** Construct the weighted-normalized matrix \( V = [v_{ij}] \) with the criteria importance coefficients: \( W=[w_j], j=1,m \), where:
  \[
  v_{ij} = w_j \cdot r_{ij}, i=1,n, j=1,m
  \]
- **Step 4.** Define the positive ideal solution vector \( V^+ \) and the negative ideal solution vector \( V^- \) (the vector for minimal value of alternatives if the criteria is a minimal criteria or the maxim value of alternatives if the criteria is a maxim criteria) thus:
  \[
  V^+ = (v_1^+, v_2^+ ... v_m^+), V^- = (v_1^-, v_2^-, ... v_m^-)
  \]
  where:
  \[
  v_{ij}^+ = \begin{cases} 
  \max_{1 \leq i \leq n} \{ v_{ij} \} & \text{when } C_j \text{ is a maximum criteria} \\
  \min_{1 \leq i \leq n} \{ v_{ij} \} & \text{when } C_j \text{ is a minimum criteria}
  \end{cases}
  \]
  and
  \[
  v_{ij}^- = \begin{cases} 
  \max_{1 \leq i \leq n} \{ v_{ij} \} & \text{when } C_j \text{ is a maximum criteria} \\
  \min_{1 \leq i \leq n} \{ v_{ij} \} & \text{when } C_j \text{ is a minimum criteria}
  \end{cases}
  \]
- **Step 5.** The calculation of distance between the i alternative and the positive ideal alternative \( V^+ \) and the calculation of distance between the i alternative and the negative ideal alternative \( V^- \):
  \[
  S_{+i} = \left[ \sum_{j=1}^{m} (v_{ij} - v_{ij}^+)^2 \right]^{1/2}, i=1,n
  \]
  \[
  S_{-i} = \left[ \sum_{j=1}^{m} (v_{ij} - v_{ij}^-)^2 \right]^{1/2}, i=1,n
  \]
- **Step 6.** The determination of the index to positive ideal solution:
  \[
  I_{Ni} = \frac{S_{-i}}{S_{+i} + S_{-i}}, i=1,n
  \]
  An alternative \( V_i \) is more near to \( V^+ \) the more \( I_{Ni} \) is close to 1.
  A number of distance metrics can be applied. TOPSIS2 is a variant where distance was
measured in least absolute value terms. Another commonly used metric is the Tchebychev metric, where the minimum maximum difference is the basis for selection [23].

**Problem of Collinearity**

We define collinearity of the criteria as a linear relationship between two criteria. Two criteria are perfectly collinear if there is an exact linear relationship between them. If X, Y are two criteria with $x_i, i=1,n$ respectively $y_i, i=1,n$ values for each alternatives, then there exist parameters $a$ and $b$ such that, for all alternatives $i$, we have [24]:

$$y_i = a + bx_i$$  \hspace{1cm} (3)

In this case the Pearson coefficient of correlation is $r=1.00$ or $r=-1.00$. (3) is the regression line equations.

We define multicollinearity as a situation in which two or more criteria are in a high linear relationship. Two or more criteria are perfectly multicollinear if there is an exact linear relationship between them. If Y is the dependent criteria and $X_j, j=1,m$ are m independent criteria, $y_i, i=1,n$ respectively $x_{ij}, i=1,n, j=1,m$ values for each alternative, then there exists parameters $a$ and $b_j, j=1,m$ such that, for all alternatives $i$, we have [24]:

$$y_i = a + b_1x_{i1} + b_2x_{i2} + \ldots + b_mx_{im}.$$ 

**Example 1.** We took data from 500 patients with diabetes mellitus type 2 recorded at Center of Diabetes, Nutrition and Metabolic Diseases, Cluj-Napoca, Romania. The TOPSIS method could be applied in this particular case for the ranking of patients in order to predict the risk for the complications of diabetes mellitus like: retinopathy, diabetic nephropathy and neuropathy. Three parameters that can be taken into account in predicting risk in this case are glycaemia, cholesterol and LDL-cholesterol. All the criteria in our example were maximum criteria. We choose these three parameters because they meet our requirements.

Mathematical model: Let be $X, Y, Z$ three criteria (glycaemia, cholesterol and LDL-cholesterol) with $x_i, i=1,500$ respectively $y_i, i=1,500$, $z_i, i=1,500$, values for each alternatives (patients with diabetes mellitus). The Pearson coefficients of correlation were $r(X,Y)=0.003, r(X,Z)=-0.03$ respectively $r(Y,Z)=0.89$.

• Phase 1. We applied TOPSIS for two independent criteria $X, Y$ (weights: $w_x=1, w_y=1$). We noted the results In$(X,Y)_i, i=1,500$.

The Pearson coefficients of correlation were $r(In(X,Y),X)=0.71$ and $r(In(X,Y),Y)=0.70$. Both of criteria have the same participation in the composition of index In.

• Phase 2. We applied TOPSIS for three criteria $X, Y$ and $Z$ ($w_x=1, w_y=1, w_z=1$). We noted the results In$(X,Y,Z)_i, i=1,500$.

The Pearson coefficients of correlation were $r(In(X,Y,Z),X)=0.59, r(In(X,Y,Z),Y)=0.78$ and $r(In(X,Y,Z),Z)=0.74$.

• Phase 3. We applied TOPSIS for two independent criteria $X, Y$ ($w_x=1, w_y=2$). We noted the results In$(X,Y)_i, i=1,500$.

The Pearson coefficients of correlation were $r(In(X,Y),X)=0.51$ and $r(In(X,Y),Y)=0.85$.

From this example we can formulate the following proposition:

**Proposition.** If $C_i, i=1,m$ criteria, where $C_k, k=1,p$ are independent criteria with $w_k, k=1,p$ weights; $C_{i'}, l=1,s$ are pairs of perfectly collinear criteria, $w_{i'}, l=1,s$ weights, $p+s=m$, then the result of TOPSIS method In$_i, i=1,n$ is the same as the results of a model with the same $C_k, k=1,p$ criteria with the weights $w_k, k=1,p$, and $C_{i'}, l=1,\frac{s}{2}$ (one of each pairs of perfectly collinear
criteria) with the weights \( w_1 = \sqrt{1 + b_1^2} \), \( l = 1 - \frac{s}{2} \), where \( b_1, l = 1 - \frac{s}{2} \) are the coefficients of regression line.

**Demonstration.** If \( s = 2 \) and \( C_i \) perfectly correlated with \( C_2 \) then, from (1) and (3), \( S_i^+ \) become:

\[
S_i^+ = \left\{ \left( \sum_{k=1}^{p} (v_{ik} - v_{i+})^2 + (v'_{i1} - v_{i+})^2 + \left[ a + b v'_1 - (a + b v'_i) \right]^2 \right)^{1/2} \right\} = \\
= \left[ \sum_{k=1}^{p} (v_{ik} - v_{i+})^2 + (v'_{i1} - v_{i+})^2 + b^2 (v'_{i1} - v_{i+})^2 \right]^{1/2} = \left[ \sum_{k=1}^{p} (v_{ik} - v_{i+})^2 + (v'_{i1} - v_{i+})^2 (1 + b^2) \right]^{1/2}
\]

Analog for \( S_i^- \).

Let \( w_k = 1, k = 1, p \) and \( w_i = t \) then:

\[
S_i^- = \left\{ \sum_{k=1}^{p} \left( w_{ik} v_{ik} - w_{i+} v_{i+} \right)^2 + \left( w_{i1} v_{i1} - w_{i+} v_{i+} \right)^2 \right\}^{1/2} = \left[ \sum_{k=1}^{p} (v_{ik} - v_{i+})^2 + t^2 (v'_{i1} - v_{i+})^2 \right]^{1/2}
\]

Analog for \( S_i^- \).

Thus \( S_i^+ = S_i^- \) if \( t = \sqrt{1 + b^2} \). Analog for \( S_i^- = S_i^- \).

**Example 2.** We took data from 304 patients with diabetes mellitus type 2 recorded at Center of Diabetes, Nutrition and Metabolic Diseases, Cluj-Napoca, Romania. For this example we use five parameters to simulate our algorithm (Patients were different than the first example). The TOPSIS method can be apply in this particular case for the ranking of patients similar with [19-20, 25]. For criteria \( C_j, j = 5, 1 \) we consider abdominal circumference, cholesterol, triglyceride, systolic blood pressure and diastolic blood pressure. All the criteria in our example were maximum criteria. We choose these parameters because they meet our requirements.

- Phase 2. We applied TOPSIS for the criteria \( C_j, j = 5, 1 \) (weights=1 for all criteria). We noted the results \( In_i, i = 1, 304 \). The Pearson coefficients of correlation were \( r(In, C_i) = 0.21, r(In, C_2) = 0.51 \), \( r(In, C_3) = 0.56, r(In, C_4) = 0.69 \) and \( r(In, C_5) = -0.71 \).

**Solution**

We propose the next algorithm:

- Step 1. Compute the Pearson correlation matrix \( r_{ni}, i, j = 1, m \) between \( C_n, j = 1, m \);
- Step 2. \( \text{max} = 0 \);
- Step 3. Get \( r_{ni} = \text{max} \) from \( r_{ni}, i, j = 1, m \); \( r_{ni}, i = j \), where \( i, j = 1, m \);
- Step 4. If \( r_{ni} \geq 0.25 \) then begin
  - \( \text{In}(C_n, C_i) = \text{Apply TOPSIS} (C_n, C_i) \);
  - Replace \( C_n \) with \( \text{In}(C_n, C_i) \);
  - Standardize \( C_n \);
  - Delete \( C_n \);
  - \( m = m - 1 \);
  - if \( m = 1 \) then Stop;
  - Go to Step 1.
- end;
- else Stop.

Where "apply TOPSIS(C_n, C_i)" means the application of TOPSIS for two criteria \( C_n \) and \( C_i \). We took the cut-off from the empirical appreciations (Colton rules [24]).
Results

We applied the algorithm in the case of the example 1. We had three criteria X, Y, Z (Cj, j=1,3) (glycaemia, cholesterol and LDL cholesterol) and 500 alternatives (patients with diabetes mellitus).

First iteration

- Step 1. The Pearson correlation matrix was presented in table 1.

Table 1. The Pearson correlation matrix r_{ij}, i=1,3, j=1,3 between criteria X, Y, Z (Cj, j=1,3)

<table>
<thead>
<tr>
<th></th>
<th>Glycaemia</th>
<th>Cholesterol</th>
<th>LDL-cholesterol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycaemia</td>
<td>1.00</td>
<td>0.003</td>
<td>-0.03</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>0.003</td>
<td>1.00</td>
<td>0.89</td>
</tr>
<tr>
<td>LDL-cholesterol</td>
<td>-0.03</td>
<td>0.89</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Step 2. max=0;
- Step 3. max=r_{32}=0.89;
- Step 4. We applied TOPSIS method for C2 and C3. We found In(C2,C3). The Pearson coefficient of correlation r(In(C2,C3),C2)=0.97 and r(In(C2,C3),C3)=0.97. We replaced C2 with In(C2,C3). We deleted C3. m=2. We repeated Step 1.

Second iteration

- Step 1. The Pearson correlation matrix was presented in Table 2.

Table 2. The Pearson correlation matrix r_{ij}, i=1,2, j=1,2 between criteria X, In(C2,C3)

<table>
<thead>
<tr>
<th></th>
<th>Glycaemia</th>
<th>In(C2,C3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycaemia</td>
<td>1.00</td>
<td>-0.02</td>
</tr>
<tr>
<td>In(C2,C3)</td>
<td>-0.02</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Step 2. max=0; Step 3. max=r_{21}=-0.02; Step 4. Stop.

The problem of the collinearity was resolved, thus TOPSIS method can be applied.

Phase 4. We applied TOPSIS for two independent criteria X, In(Y,Z) (w_x=1, w_m=1). We noted the results In(X,In), i=1,500.

The Pearson coefficients of correlation were r(In(X,In),X)=0.73, r(In(X,In),Y)=0.66 and r(In(X,In),Z)=0.64.

We applied the algorithm in the case of second example. We had five criteria Cj, j=1,5 (abdominal circumference, cholesterol, triglyceride, systolic blood pressure and diastolic blood pressure) and 304 alternatives (patients with diabetes mellitus).

First iteration

- Step 1. The Pearson correlation matrix was presented in Table 3.

Table 3. The Pearson correlation matrix r_{ij}, i=1,5, j=1,5 between criteria Cj, j=1,5

<table>
<thead>
<tr>
<th></th>
<th>Abdominal circumference</th>
<th>Cholesterol</th>
<th>Triglyceride</th>
<th>Systolic blood pressure</th>
<th>Diastolic blood pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal circumference</td>
<td>1.00</td>
<td>0.24</td>
<td>-0.08</td>
<td>-0.09</td>
<td>-0.10</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>0.24</td>
<td>1.00</td>
<td>0.30</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Triglyceride</td>
<td>-0.08</td>
<td>0.30</td>
<td>1.00</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>-0.09</td>
<td>0.01</td>
<td>0.16</td>
<td>1.00</td>
<td>0.62</td>
</tr>
<tr>
<td>Diastolic blood pressure</td>
<td>-0.10</td>
<td>-0.01</td>
<td>0.16</td>
<td>0.62</td>
<td>1.00</td>
</tr>
</tbody>
</table>
• Step 2. \( \text{max}=0; \) Step 3. \( \text{max}=r_{54}=0.62; \)
• Step 4. We applied TOPSIS method for \( C_3 \) and \( C_4 \). We found \( \text{In}(C_4, C_5) \). The Pearson coefficient of correlation \( r(\text{In}(C_4, C_5), C_4)=0.87 \) and \( r(\text{In}(C_4, C_5), C_5)=0.92 \). We replaced \( C_4 \) with \( \text{In}(C_4, C_5) \). We deleted \( C_5 \). \( m=4 \). We repeated Step 1.

**Second iteration**
• Step 1. The Pearson correlation matrix was presented in Table 4.

**Table 4.** The Pearson correlation matrix \( r_{ij}, i=1, 4, j=1, 4 \) between criteria \( C_1, C_2, C_3, \text{In}(C_4, C_5) \)

<table>
<thead>
<tr>
<th></th>
<th>Abdominal circumference</th>
<th>Cholesterol</th>
<th>Triglyceride</th>
<th>( \text{In}(C_4, C_5) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal circumference</td>
<td>1.00</td>
<td>0.24</td>
<td>-0.08</td>
<td>-0.10</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>0.24</td>
<td>1.00</td>
<td>0.30</td>
<td>0.00</td>
</tr>
<tr>
<td>Triglyceride</td>
<td>-0.08</td>
<td>0.30</td>
<td>1.00</td>
<td>0.18</td>
</tr>
<tr>
<td>( \text{In}(C_4, C_5) )</td>
<td>-0.10</td>
<td>0.00</td>
<td>0.18</td>
<td>1.00</td>
</tr>
</tbody>
</table>

• Step 2. \( \text{max}=0; \) Step 3. \( \text{max}=r_{32}=0.30; \)
• Step 4. We applied TOPSIS method for \( C_2 \) and \( C_3 \). We found \( \text{In}(C_2, C_3) \). The Pearson coefficient of correlation \( r(\text{In}(C_2, C_3), C_2)=0.86 \) and \( r(\text{In}(C_2, C_3), C_3)=0.89 \). We replaced \( C_2 \) with \( \text{In}(C_2, C_3) \). We deleted \( C_3 \). \( m=3 \). We repeated Step 1.

**Third iteration**
• Step 1. The Pearson correlation matrix was presented in Table 5.

**Table 5.** The Pearson correlation matrix \( r_{ij}, i=1, 3, j=1, 3 \) between criteria \( C_1, \text{In}(C_2, C_3), \text{In}(C_4, C_5) \)

<table>
<thead>
<tr>
<th></th>
<th>Abdominal circumference</th>
<th>( \text{In}(C_2, C_3) )</th>
<th>( \text{In}(C_4, C_5) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal circumference</td>
<td>1.00</td>
<td>0.08</td>
<td>-0.10</td>
</tr>
<tr>
<td>( \text{In}(C_2, C_3) )</td>
<td>0.08</td>
<td>1.00</td>
<td>0.13</td>
</tr>
<tr>
<td>( \text{In}(C_4, C_5) )</td>
<td>-0.10</td>
<td>0.13</td>
<td>1.00</td>
</tr>
</tbody>
</table>

• Step 2. \( \text{max}=0; \)
• Step 3. \( \text{max}=r_{21}=0.13; \)
• Step 4. Stop

The problem of the collinearity was resolved, thus TOPSIS method can be applied.

Phase 4 (for example 2). We applied TOPSIS for three independent criteria \( C_1, \text{In}(C_2, C_3), \text{In}(C_4, C_5) \) (weights=1 for all criteria). We noted the results \( \text{In}_i, i=1, 304.1 \). The Pearson coefficients of correlation were \( r(\text{In}, C_1)=0.58, r(\text{In}, C_2)=0.63, r(\text{In}, C_3)=0.55, r(\text{In}, C_4)=0.42 \) and \( r(\text{In}, C_5)=0.42 \).

**Discussion**

In example 1 criterion \( X \) was not correlated with the other two criteria \( Y \) and \( Z \), but \( Y \) was strong correlated with \( Z \). The index result from TOPSIS (Phase4) was more correlated with \( Y \) \( r(\text{In}(X, Y, Z), Y)=0.78 \) and \( Z \) \( r(\text{In}(X, Y, Z), Z)=0.74 \), than with \( X \) \( r(\text{In}(X, Y, Z), X)=0.59 \). Even if \( X \) and \( Y \) were not perfectly correlated they influenced the results.

In example 2 criterion \( C_1 \) was not correlated with the other four criteria. The index result from TOPSIS (Phase4) was more correlated with \( C_2-C_5 \) \( r(\text{In}(C_2), 0.51, r(\text{In}(C_3)=0.56, r(\text{In}(C_4)=0.69 \) and \( r(\text{In}(C_5)=-0.71 \), than with \( C_1 \) \( r(\text{In}, C_1)=0.21 \).

After we applied the purposed algorithm the index result from TOPSIS (Phase4) was equal correlated with all the criteria (for first example \( r(\text{In}(X, \text{In}), X)=0.73, r(\text{In}(X, \text{In}), Y)=0.66 \) and...
Correlated Criteria in Decision Models: Recurrent Application of TOPSIS Method

In our study, we applied TOPSIS method in a recurrent manner until the coefficients of correlation were less than 0.25. The disadvantage of our algorithm is that we took the cut-off from the empirical appreciations (Colton rules [24]) and is not necessary that a correlation less than 0.25 not influence the results of TOPSIS method. This cut-off 0.25 is debatable. It remained for further study to establish a better way to end the algorithm.

The algorithm can be used for weights different than 1.

The proposed algorithm has a strong empirical justification relying on examples. It remained for further study to compare the algorithm proposed in this paper with different methods.

### Conclusion

The proposed algorithm can solve the problem of multicollinearity between criteria in the problem of decision making. The purpose of our algorithm was to apply recurrent the TOPSIS method until the correlations between components are reduced below a threshold.

### List of abbreviations

TOPSIS = Technique for Order Preference by Similarity to Ideal Solution

### Conflict of Interest

The authors declare that they have no conflict of interest.

### Acknowledgements

The database for examples 1 and 2 has been gratefully obtained with the help of medical staff from Center of Diabetes, Nutrition and Metabolic Diseases, Cluj-Napoca, Romania.

The author CB was partly supported by European Social Fund within the Sectorial Operational Program - Human Resources Development 2007-2013-POSDRU /89/1.5/S/58965.

### References


