An Extended Case-Based Distance Approach for Alternatives Screening

Li-Ching Ma\textsuperscript{1,*}, Pei-Pei Hsu\textsuperscript{2}
\textsuperscript{1}Department of Information Management, National United University
No.1, Lienda, MiaoLi City, Taiwan 36003, R.O.C.
\textsuperscript{2} Department of Mechanical and Automation Engineering, I-Shou University
No.1, Sec. 1, Syuecheng Rd., Dashu Township, Kaohsiung County 840, Taiwan, R.O.C.
\textsuperscript{1}lcma@nuu.edu.tw \textsuperscript{2}pphsu@isu.edu.tw

Abstract

Screening is a helpful process to reduce larger set of alternatives into a smaller set that contains the best alternatives, thus decision makers can concentrate on evaluating alternatives in the smaller set. Hence, how to assist decision makers in screening out poor alternatives is an important issue in multiple criteria decision making. This study tries to develop a screening model incorporating the advantages of the case-based distance method and the discriminant analysis. From the concept of the case-based distance method, the proposed approach can obtain criterion weights and screening rules by eliciting decision makers’ preferences based on a set of test cases. Moreover, the proposed approach can increase hit rates and improve multiple solution problems of conventional case ed-based distance methods.

Keywords: Decision-making, Screening, Case, Distance, Discriminant analysis

1. Introduction

The major tasks of multiple-criteria decision aid (MCDA) are to assist a decision maker to choose, rank, or sort alternatives of a finite set (Roy, 1996). Screening is a process to reduce larger set of alternatives into a smaller set that contains the best alternatives, thus decision makers can concentrate on evaluating alternatives in the smaller set. Therefore, how to aid decision makers in screening out poor alternatives is an important issue in multiple-criteria decision making.

Chen et al. (2008) proposed a case-based distance method for screening in MCDA.
Case-based reasoning is to get preferential information by evaluating a test set selected by the decision maker. The weights of criteria and a distance threshold are obtained from the test set, both of which are used to filter out less preferred alternatives. This screening method based on previous cases and distance measure is simple and easy to understand. However, the quadratic optimization model used to screen out alternatives may result in lower hit rates and multiple solution problems.

Discriminant Analysis (DA) is a statistical technique for predicting group membership, which can also be used in screening alternatives. The GP (Goal Programming)-based DA, first proposed by Freed and Glover (1981), can estimate weights of criteria by minimizing sum of deviations (M, SD, Freed and Glover, 1986) or minimizing misclassified observations (MMO, Banks and Abad, 1991). Those weights yield an evaluation score, which is compared with a threshold value for classifying observations. In order to improve hit rates, Sueyoshi (2004) and Sueyoshi and Hwang (2004) developed a mixed integer programming (DA-MIP) approach to estimate weights by minimizing the total number of misclassified observations. This approach can efficiently improve hit rates; however, it may cause multiple solution problems.

This study tries to develop an extended case-based distance approach to assist decision makers in screening alternatives. By incorporating the advantages of the DA-MIP method (Sueyoshi, 2004; Sueyoshi and Hwang, 2004) and the case-based distance method (Chen, 2008), the proposed approach can screen alternatives by evaluating a test set of cases selected by the decision maker, increase hit rates and improve multiple solution problems.

2. The Case-Based Distance Method

This section briefly reviews the case-based distance method (Chen et al., 2008). Given a set of alternatives \( A = \{A_1, A_2, \ldots, A_i, \ldots, A_n\} \), which are compared with respect to each of the \( q \) criteria \( C_1, \ldots, C_q \). The values of criteria \( C_1, \ldots, C_q \) for an alternative \( A_i \) are expressed as \( c_{ij} \), for \( j = 1, \ldots, q \). Suppose that a decision maker specifies \( t \) selected cases as \( T = \{Z_1, Z_2, \ldots, Z_r, \ldots, Z_t\} \) in a choice problem. Let \( Z = \{Z_1, Z_2, \ldots, Z_m\} \) denote the \( m \) acceptable cases and \( T - Z = \{Z_{m+1}, Z_{m+2}, \ldots, Z_t\} \) denote the \( t-m \) unacceptable cases.

Denote \( Z_* \) as a fictitious alternative at the center of \( Z \), where

\[
c_{*j} = \frac{1}{m} \sum_{r=1}^{m} c_{rj}.
\]  

(2.1)
The distance between \( Z_r \in \mathbf{Z} \) and the center \( Z_* \) on criterion \( j \) is defined as

\[
d_j(Z_r, Z_*) = \frac{(c_{r,j} - c_{*,j})^2}{d_j^{\text{max}}}. \tag{2.2}
\]

The denominator \( d_j^{\text{max}} \) is the maximum of \( \{(c_{r,j} - c_{*,j})^2\} \) for \( r = 1, \ldots, m \), which is used as a normalization factor for criterion \( j \). The distance between a case \( Z_r \) and the target case \( Z_* \) is defined as follows:

\[
D(Z_r, Z_*) = \sum_{j=1}^{q} w_j \times d_j(Z_r, Z_*), \tag{2.3}
\]

where \( w_j \) represents the weight of criterion \( j \). Let \( R \) be the distance threshold to distinguish cases in \( \mathbf{Z} \) from cases in \( \mathbf{T} \setminus \mathbf{Z} \).

Let \( \alpha_r \) and \( \alpha_p \) be error adjustment parameters or the DM’s inconsistent judgments on \( \mathbf{Z} \) and \( \mathbf{T} \setminus \mathbf{Z} \) respectively.

\[
D(Z_r, Z_*) + \alpha_r \leq R, \quad -1 \leq \alpha_r \leq 0 \quad \text{for} \quad r = 1, \ldots, m, \tag{2.4}
\]

\[
D(Z_r, Z_*) + \alpha_p \geq R, \quad 0 \leq \alpha_p \leq 1 \quad \text{for} \quad r, p = m+1, \ldots, t. \tag{2.5}
\]

The case-based distance model was constructed as follows:

**The Case-Based Distance Model (Chen et al., 2008)**

Min \( ERR = \sum_{r=1}^{m} (\alpha_r)^2 + \sum_{p=m+1}^{t} (\alpha_p)^2 \)

s.t. \( \sum_{j=1}^{q} w_j \times d_j(Z_r, Z_*) + \alpha_r \leq R, \quad \forall \ r = 1, 2, \ldots, m, \tag{2.6} \)

\( \sum_{j=1}^{q} w_j \times d_j(Z_r, Z_*) + \alpha_p \geq R, \quad \forall \ r, p = m+1, \ldots, t, \tag{2.7} \)

\( 0 < R \leq 1, \tag{2.8} \)

\( w_j > 0, \quad \forall \ j = 1, 2, \ldots, q, \tag{2.9} \)
\[
\sum_{j=1}^{d} w_j = 1, \tag{2.10}
\]

\[-1 \leq \alpha_r \leq 0, \quad \forall r = 1, 2, \ldots, m, \tag{2.11}\]

\[0 \leq \alpha_p \leq 1, \quad \forall p = m+1, \ldots, t, \tag{2.12}\]

Let \( w^* = (w_1^*, w_2^*, \ldots, w_d^*) \) and \( R^* \) be the optimal solution obtained from the above case-based distance model. An alternative \( A_i \) can be screened by the following rules: if \( D(A_i, Z_*) \leq R^* \), then \( A_i \) should be retained; otherwise, \( A_i \) should be screened out.

The objective of the case-based distance model is to minimize overall squared errors. However, the number of incorrect classifications may be large. In addition, when the decision maker’s judgments are consistent (i.e. \( \alpha_r = \alpha_p = 0 \)), the model may yield multiple solutions which may result in unstable screening results.

3. The Proposed Extended Case Based Distance Approach

This study tries to develop an extended case based distance approach to assist decision makers in screening alternatives. Given a decision problem with \( n \) alternatives \( A = \{A_1, A_2, \ldots, A_n\} \). Suppose that the decision maker specifies \( t \) selected cases as \( T = \{Z_1, Z_2, \ldots, Z_r, \ldots, Z_t\} \). Let \( Z = \{Z_1, Z_2, \ldots, Z_m\} \) denote the \( m \) acceptable cases and \( T - Z = \{Z_{m+1}, Z_{m+2}, \ldots, Z_t\} \) denote the \( t - m \) unacceptable cases. Let \( Z_* \) be the target case. \( Z_* \) can be selected as an ideal case with each criterion value \( c_{*,j} \) equal to the best value of creation \( j \) for all acceptable cases in \( Z \) or a center case with \( c_{*,j} = \frac{1}{m} \sum_{r=1}^{m} c_{r,j} \).

In order to improve hit rates, this study develops an optimization model to obtain the most appropriate weight set \( w^* \) and distance threshold \( R^* \) by minimizing the number of incorrect classifications instead of minimizing errors caused by the decision maker’s inconsistent judgments. The proposed extended case-based distance model is constructed as follows:
Model 1

Min \quad M \times Obj1 + Obj2 \tag{3.1}

\begin{align*}
Obj1 &= \sum_{r=1}^{m} h_r + \sum_{r=m+1}^{t} h_r \tag{3.2} \\
Obj2 &= R \tag{3.3}
\end{align*}

s.t. \quad \sum_{j=1}^{q} w_j \left( \frac{c_{r,j} - c_{*,j}}{d_{j}^{\max}} \right)^2 - M h_r \leq R, \quad \forall r = 1, ..., m, \tag{3.4}

\sum_{j=1}^{q} w_j \left( \frac{c_{r,j} - c_{*,j}}{d_{j}^{\max}} \right)^2 + M h_r \geq R + \varepsilon, \quad \forall r = m + 1, ..., t, \tag{3.5}

\sum_{j=1}^{q} w_j = 1, \tag{3.6}

0 \leq R \leq 1, \tag{3.7}

h_r \in \{0,1\}, \quad \forall r = 1, ..., t, \quad M \text{ is a large value, } \varepsilon \text{ is a tolerance error.} \tag{3.8}

The first objective (Obj1) is to minimize the number of incorrect classifications by counting a binary variable \( h_r \) for \( r = 1, ..., t \). The second objective (Obj2) is used to improve the multiple solution problems. That is, when multiple solutions achieve the same optimal objective (Obj1), the one yielding minimum \( R \) will be chosen as the only optimal solution. Because minimizing the incorrect classifications is the most important objective, Obj1 is multiplied by a big value \( M \) (for example, \( M = 10^6 \)). Expression (3.4) is used to classify all acceptable cases in \( Z \). If an acceptable case \( Z_r \) is correctly classified (i.e., \( \sum_{j=1}^{q} w_j \left( \frac{c_{r,j} - c_{*,j}}{d_{j}^{\max}} \right)^2 \leq R \) then \( h_r = 0 \); otherwise \( h_r = 1 \). Similarly, Expression (3.5) is used to classify all unacceptable cases in \( T - Z \). Expressions (3.6) sets the sum of weights equal to one. Since all distance values have been normalized, the range of distance threshold \( R \) is set between 0 and 1, as expressed in (3.7).

A new observation \( A_i \) can be screened by the following rules: If

\[ \sum_{j=1}^{q} w_j \left( \frac{c_{i,j} - c_{*,j}}{d_{j}^{\max}} \right)^2 \leq R^*, \quad \text{then } A_i \text{ is in the acceptable set; otherwise } A_i \text{ is in the unacceptable} \]
set. Model 1 can be solved directly by global optimization software. For instance, global solver of Lingo 9.0 is adopted in this study.

4. A Numerical Example

The following example, modified from Harvard Business Review (Hammond et al., 1998), is applied to illustrate the proposed approach. The example describes a business problem for renting an office. A decision maker defines five major objectives to fulfill in selecting his/her office: (i) a short commute time from home to office, (ii) good access to his clients, (iii) good office service, (iv) sufficient space, and (v) low costs. The commuting time is the average time in minutes needed to travel to work during rush hour. The percentage of his clients within an hour’s drive to the office is used to measure the access to clients. Office service is represented by a value between 0 and 10 with the best value 10. Office size is measured in square feet, and cost is measured by monthly rent. The decision maker hopes to keep monthly cost and commuting time as small as possible and remaining criteria larger.

The test set $\mathbf{T} = \{ Z_1, Z_2, \ldots, Z_7 \}$ shown in Table 1 is presented to the decision maker.

Assume the decision maker selects $\mathbf{Z} = \{ Z_1, Z_2, Z_3 \}$ as acceptable cases and $\mathbf{T} - \mathbf{Z} = \{ Z_4, Z_5, Z_6, Z_7 \}$ as unacceptable cases. Assume the decision maker specifies the upper bound for each criterion as listed in the last column of Table 1 and selects the target case $Z_s$ as an ideal case. In order to keep all weights positive, monthly cost and commute time are transformed to the positive format by subtracting from the upper bound.

Applying the case-based distance model (Chen et al., 2008) to the example yields $\mathbf{w}^* = (0.2, 0.01, 0.01, 0.36, 0.42)$, $R^* = 0.595$, and $ERR = 0.0008$. The distances of the cases to the target are listed in Table 2. The four cases $Z_1, Z_3, Z_4$, and $Z_7$ are misclassified. Although the overall squared error in $\mathbf{T}$ is low, the rate of misclassification for test cases is very high (4/7 = 57%). In order to make comparisons, applying the proposed model to the same example yields $\mathbf{w}^* = (0.30, 0.49, 0, 0, 0.22)$, $R^* = 0.336$, $Obj1 = 1$, and $Obj2 = 0.336$. The binary variable $h_3$ is equal to 1 and the remaining $h_r$ is equal to 0 for all $r \neq 3$. The distances of the cases to the target by the proposed approach are listed in Table 3, where only one test case $Z_3$ is misclassified. The rate of misclassification for test cases is 1/7 = 14%. The tolerance error $\varepsilon$ is set as 0.001 in this study. Since the value of $R^*$ is obtained, all alternatives can be screened by the following rule: if $D(A_i, Z_s) \leq R^*$, then $A_i$ should be
retained; otherwise, \( A_i \) should be screened out.

As shown in the office renting example, when a decision maker's judgments are inconsistent, the proposed approach yields higher hit rates than that of the original case based distance method. On the other hand, when a decision maker's judgments are consistent, the original case-based distance method may result in multiple solutions because there could be many \( R^* \)'s achieve the same minimum objective value (\( ERR = 0 \)). The proposed approach tries to improve the multiple solution problems of both the original case-based distance method and DA-MIP method by adding a second objective function (\( Obj_2 \)). When there are more than one \( R^* \)'s achieve the same optimal \( Obj_1 \), the minimum one will be adopted. Therefore, a unique solution \( R^* \) can be obtained.

5. Conclusions

This study tries to assist decision makers in solving screening problems in multiple criteria decision making. From the concept of the case-based distance method (Chen, 2008), the proposed approach can obtain criterion weights and screening rules by eliciting decision makers’ preferences based on a set of test cases. The proposed approach can increase hit rates by incorporating the advantages of the DA–MIP method (Sueyoshi, 2004; Sueyoshi and Hwang, 2004). Moreover, the multiple solution problems of both original case-based distance method and DA-MIP method can also be improved.

Future research could address the issues about how to provide visual aids to assist decision makers in observing background information and screening out poor alternatives.

References


Table 1  The test set of the office-renting example

<table>
<thead>
<tr>
<th></th>
<th>$Z_1$</th>
<th>$Z_2$</th>
<th>$Z_3$</th>
<th>$Z_4$</th>
<th>$Z_5$</th>
<th>$Z_6$</th>
<th>$Z_7$</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>Commute (Mins)</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>25</td>
<td>45</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>$C_2$</td>
<td>Customer Access (%)</td>
<td>85</td>
<td>90</td>
<td>75</td>
<td>90</td>
<td>65</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>$C_3$</td>
<td>Office Services</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$C_4$</td>
<td>Office Size (Square Feet)</td>
<td>800</td>
<td>950</td>
<td>900</td>
<td>850</td>
<td>750</td>
<td>700</td>
<td>900</td>
</tr>
<tr>
<td>$C_5$</td>
<td>Monthly Cost ($)</td>
<td>1650</td>
<td>1700</td>
<td>1800</td>
<td>1900</td>
<td>1800</td>
<td>1950</td>
<td>1700</td>
</tr>
</tbody>
</table>

Table 2  Distances of the cases to the target by the case-based distance method (Chen et al., 2008)

<table>
<thead>
<tr>
<th></th>
<th>$Z_1$</th>
<th>$Z_2$</th>
<th>$Z_3$</th>
<th>$Z_4$</th>
<th>$Z_5$</th>
<th>$Z_6$</th>
<th>$Z_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>Commute</td>
<td>0.062</td>
<td>0.090</td>
<td>0.122</td>
<td>0.062</td>
<td>0.202</td>
<td>0.090</td>
</tr>
<tr>
<td>$C_2$</td>
<td>Customer</td>
<td>0.002</td>
<td>0.001</td>
<td>0.005</td>
<td>0.001</td>
<td>0.010</td>
<td>0.007</td>
</tr>
<tr>
<td>$C_3$</td>
<td>Services</td>
<td>0.004</td>
<td>0.001</td>
<td>0.004</td>
<td>0.001</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>$C_4$</td>
<td>Size</td>
<td>0.231</td>
<td>0.090</td>
<td>0.130</td>
<td>0.130</td>
<td>0.293</td>
<td>0.362</td>
</tr>
<tr>
<td>$C_5$</td>
<td>Cost</td>
<td>0.298</td>
<td>0.316</td>
<td>0.355</td>
<td>0.395</td>
<td>0.355</td>
<td>0.416</td>
</tr>
<tr>
<td>Distance</td>
<td>0.598</td>
<td>0.499</td>
<td>0.617</td>
<td>0.590</td>
<td>0.870</td>
<td>0.885</td>
<td>0.577</td>
</tr>
</tbody>
</table>

Table 3  Distances of the cases to the target by the proposed approach

<table>
<thead>
<tr>
<th></th>
<th>$Z_1$</th>
<th>$Z_2$</th>
<th>$Z_3$</th>
<th>$Z_4$</th>
<th>$Z_5$</th>
<th>$Z_6$</th>
<th>$Z_7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_1$</td>
<td>Commute</td>
<td>0.091</td>
<td>0.131</td>
<td>0.179</td>
<td>0.091</td>
<td>0.296</td>
<td>0.131</td>
</tr>
<tr>
<td>$C_2$</td>
<td>Access</td>
<td>0.089</td>
<td>0.040</td>
<td>0.249</td>
<td>0.040</td>
<td>0.487</td>
<td>0.358</td>
</tr>
<tr>
<td>$C_3$</td>
<td>Services</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$C_4$</td>
<td>Size</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$C_5$</td>
<td>Cost</td>
<td>0.156</td>
<td>0.165</td>
<td>0.185</td>
<td>0.206</td>
<td>0.185</td>
<td>0.217</td>
</tr>
<tr>
<td>Distance</td>
<td>0.336</td>
<td>0.336</td>
<td>0.612</td>
<td>0.337</td>
<td>0.968</td>
<td>0.707</td>
<td>0.503</td>
</tr>
</tbody>
</table>