Methods for Choosing Shortlists of Potentially Attractive Offers from a Large Data Set

David Ramsey[1]
Department of Operations Research
Wrocław University of Science and Technology, Poland

[1] The speaker acknowledges the co-authorship of Aleksander Mariański, Leopold Szczurowski and Michał Kędziora

Introduction

When looking for a unique, valuable resource (e.g. a flat), decision makers (DMs) can find basic information about a wide range of offers via the Internet.

This information can be used to direct search, e.g. using the shortlist heuristic.

Based on initial information from the Internet, DMs select a shortlist of potentially attractive offers that are to be physically viewed.

After viewing the offers on the shortlist, DMs make their final selection.

This talk considers methods for constructing such a shortlist.

Assumptions Regarding the Shortlist

- Initial Shortlists are created using quantitative information about offers available via the Internet.
- A Shortlist of offers to view should include offers that are i) all potentially highly attractive, ii) relatively diverse.
- It should be noted that since qualitative information (e.g. photos) may also be available via the Internet, the method presented here can also be interpreted as an initial filter in constructing a shortlist.

Example: The Data Used for Selecting a Shortlist of Flats to View

Suppose that the following quantitative data are available (possibly indirectly) from the Internet.

- 1. Price (in PLN), X_1.
- 2. Size (in m^2), X_2.
- 3. No. of rooms, N.
- 4. Location (given by two coordinates relative to a reference point), (X_3, X_4).

It should be noted that the location of an offer can be used to define its distance from a desirable location (e.g. city centre, workplace).

Hard Constraints

The DM can define hard constraints that an offer has to satisfy, e.g. a minimum and maximum number of rooms, maximum price.

Offers that do not satisfy these constraints are not considered.

It should be noted that if the number of offers satisfying the hard constraints is deemed to be too small, then these constraints can be relaxed.

After this, a vector describing the potential attractiveness of each of the remaining offers and a matrix describing the difference (distance) between offers are defined on the basis of the data available.

Assessment of the Attractiveness of Offers

The attractiveness of offers can be assessed using any appropriate method for multi-criteria decision making (e.g. TOPSIS, simple additive weighting - SAW).

Here, only continuous traits are considered at this stage (i.e. the number of rooms was only used in the process of initial filtering based on the hard constraints).

For example, the attractiveness of an offer can be assessed according to i) price X_1, ii) size, X_2 iii) distance from the city center, $U = (X_3^2 + X_4^2)^{0.5}$.

These three variables can be scaled linearly to take values in the interval $[0, w_i]$, where w_i is the weight of the i-th trait.

Assessment of the Attractiveness of Offers: TOPSIS

Using TOPSIS price and distance are assumed to be disincentives and size is assumed to be an incentive.

The attractiveness of an offer is given by $A = \frac{D^-}{D^- + D^+}$, where

- 1. D^-: Distance of scaled vector (x_1, x_2, u) from the anti-ideal, $(w_1, 0, w_3)$.
- 2. D^+: Distance of scaled vector (x_1, x_2, u) from the ideal, $(0, w_2, 0)$.

Assessment of the Attractiveness of Offers: SAW

For comparison with the TOPSIS method, it is assumed that price and distance are disincentives and size is an incentive.

Hence, using linear weighting: $A = (w_1 - x_1) + x_2 + (w_3 - u)$.

Note that the linear weighting approach can be relatively easily adapted to the assumption that an intermediate value of a variable, e.g. size, is optimal.

Using either method, the attractiveness scores belong to the interval [0,1].

Assessment of the Diversity of Offers

Assume that the diversity of offers can be assessed according to i) price X_1, ii) size, X_2 iii) location, (X_3, X_4).

Each of these variables is linearly scaled to the inverval [0,1].

The distance between offer i and offer j is defined to be

$$d_{i,j} = \sqrt{(y_{1,i} - y_{1,j})^2 + (y_{2,i} - y_{2,j})^2 + (y_{3,i} - y_{3,j})^2 + (y_{4,i} - y_{4,j})^2},$$

where $y_{m,i}$ is the scaled value of the m-th variable for the i-th offer

Goal of the Algorithm

The goal of the algorithm is to construct a shortlist of offers that maximizes a weighted average of a) the mean attractiveness of the offers on the shortlist and b) the diversity of the offers on the shortlist (measured by the mean distance between the offers).

$$V = (1 - \alpha)\overline{A} + \alpha\overline{D},$$

where $\overline{A}$ and $\overline{D}$ are the mean attractiveness of offers on the shortlist and the mean distance between offers on the shortlist, respectively.

Hence, α is the weight given to diversity.

Goal of the Algorithm

Suppose that n offers satisfy the hard constraints and we wish to construct a shortlist of k offers.

There are ${}_nC_k$ possible shortlists.

When n is large, for practical values of k (say, 10-100), exhaustive evaluation of all the possible shortlists may well be impractical for online use.

Hence, we propose a greedy algorithm that constructs a near optimal shortlist based on the objective function.

Algorithm

- 1. The number of items to place on the shortlist, k, is given.
- 2. The offer assessed to be most attractive is placed on the shortlist.
- 3. At each successive stage, we add the offer not yet on the list which maximizes the "value" of the shortlist, V, until k offers have been added.

Algorithm

When $\alpha = 0$ (no weight is given to diversity), this algorithm simply places the *k* most highly ranked offers on the shortlist.

In this case, the algorithm always finds the "optimal" shortlist (from the principles of dynamic programming).

For $\alpha = 1$, the algorithm attempts to maximize the diversity of the offers under the constraint that the highest ranked offer is placed on the shortlist.

In this case, the value of the final shortlist cannot be expressed as a sum of the "values" of its component parts.

Implementation of the Algorithm

11 321 offers of properties in the city of Wrocław from www.otodom.pl (obtained on 4/2/2021).

Location is not considered as a factor.

Hard constraints: i) Price $<$ 500 000, ii) Size $> 50m^2$, iii) no. of rooms between 3 and 5.

917 offers satisfied these hard constraints.

Implementation of the Algorithm

The assessment of the attractiveness of an offer is based on price and size, which were ascribed weights of 0.3 and 0.7, respecitvely.

The weight ascribed to the diversity of offers $\alpha \in \{0, 0.1, 0.2, \ldots, 1\}$.

Shortlists of length 30 were constructed using either TOPSIS or SAW to assess the attractiveness of an offer.

We consider the relative attractiveness of the offers placed on the short list (compared to the shortlist formed when $\alpha = 0$) and the relative diversity of these offers (compared to the shortlist formed when $\alpha = 1$).

Results

When $\alpha = 0$, the list constructed using SAW includes all of the offers selected via TOPSIS except for the 26th ranked offer (according to TOPSIS).

When $\alpha = 1$, the list constructed using SAW is the same as the list constructed using TOPSIS (since the two methods select the same offer to be the highest ranked).

For intermediate values of α (≈ 0.5), the relative attractiveness of the offers on the short list are high ($\geqslant 0.98$), while the diversity of offers is significantly increased compared to the case where $\alpha = 0$.

In these cases, the 15 most highly ranked offers (according to either procedure) are placed on the shortlist.

Conclusion

The algorithm presented here can be used to construct shortlists of offers from a large database that are simultaneously a) potentially highly attractive and b) relatively diverse.

The algorithm can be adapted to a range of methods for assessing the attractiveness of offers based on multiple traits.

There are very few differences between the results obtained using TOPSIS and the results obtained using SAW.

SAW is more flexible in practical applications.

Future Research

The results indicate that the shortlists constructed via this algorithm have desirable theoretical properties.

Questions remain as to the practical use of this algorithm:

- 1. Do DMs find such a procedure a) simple to use, b) effective?
- 2. What weight should be given to the diversity measure?
- a. This should depend on the knowledge of the DM regarding the market.
- b. The range of the attractiveness measures is independent of the number of variables used, however the range of the distance measures is not. Hence, it might be necessary to adapt the weight ascribed to diversity to the number of variables used to define the distance measure.