2022 Volume XXII(1): 109-120
DOI: 10.25142/aak.2022.009

ON CREATING SHORTLISTS FROM A LARGE DATABASE OF OFFERS

[Vytváření užších seznamů z rozsáhlé databáze nabídek]

David Ramsey[1], **Aleksander Mariański**[2], **Leopold Szczurowski**[3], **Michał Kędziora**[4]

[1] Wrocław University of Science and Technology, Faculty of Computer Science and Telecommunications, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
Email: david.ramsey@pwr.edu.pl

[2] Wrocław University of Science and Technology, Faculty of Computer Science and Telecommunications, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
Email: aleksander.marianski@pwr.edu.pl

[3] Wrocław University of Science and Technology, Faculty of Management, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
Email:leopold.szczurowski@pwr.edu.pl

[4] Wrocław University of Science and Technology, Faculty of Computer Science and Telecommunications, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
Email: michal.kedziora@pwr.edu.pl

Abstract: Nowadays, consumers can find basic information about a wide range of offers with almost no effort. For example, Internet sites give information about second-hand cars regarding the mark, age, price, engine type/size and distance travelled. This information suffices to assess whether a car is potentially attractive, but is not sufficient to make a final decision. The authors have developed an automatic procedure for selecting a shortlist of offers that aims to maximize a weighted average of the attractiveness and diversity of the offers on the shortlist. The consumer determines the number of offers to be placed on the shortlist and the relative importance of the traits considered. The authors have compared the results obtained by applying two methods of multicriteria assessment.

Keywords: consumer decision, heuristic algorithm, job search problem, multi-criteria decision process.

JEL classification: C44, C61, D11, D83

Received: 31.1.2021; Reviewed: 4.3.2021; 17.3.2021; Accepted: 18.5.2022

Introduction

Due to the rise of information technology, consumers can find information regarding a very large number of products with almost no effort. Since humans' ability to process information is limited, this may lead to the so-called paradox of choice, see Schwartz (2004). This paradox lies in the fact that when there exist a large number of offers, a decision maker (DM) might gain by only paying attention to a small set of offers rather than expending a large amount of effort with the aim of selecting the best of the offers. A DM who concentrates his/her efforts on a small subset of offers will almost certainly not purchase the best offer according to the DM's preferences. However, since the DM has invested less time and effort in making a purchase, the expected overall reward from such a search procedure is likely to be greater.

Due to limits on human cognition, DMs adopt heuristic rules that are suited both to the structure of the information available and the abilities of the DM (see Simon 1956, Todd and Gigerenzer 2000). Assume that a DM can gain some useful information about offers very cheaply, but relatively close inspection is required to accurately assess an offer. This will be the case, for example, when an individual is looking for a flat or a second-hand car. Basic information can be easily found via the Internet. However, physically viewing an offer is necessary to accurately appraise its value. To reduce search costs, a DM can select a shortlist of offers to physically view based on information from the Internet. Very often, Internet sites give a standardized set of traits describing an offer. Many of these traits are either numerical or fall into clear categories, e.g. type of engine. Hence, automatic procedures enabling the selection of shortlists of offers from databases may be of great practical use.

It is assumed that the goal of creating a shortlist is to ultimately select a single valuable resource, instead of using a shortlist to choose multiple options (e.g. Durbach and Davis 2012, present a model of choosing several ways of reducing electricity bills). We concentrate on the initial step of the selection procedure, i.e. choosing the shortlist. On the other hand, Belton (1985) considers how a DM should make a final choice from an initial shortlist.

Several recent articles have considered the concept and theoretical properties of shortlists (e.g. Masatlioglu et al. 2012, Lleras et al. 2017). Borah and Kops (2019) model search processes under which decision makers form shortlists on the basis of information from peers. The shortlist heuristic can be applied when offers are described by multiple criteria (Dulleck et al. 2011) or can be placed in clear categories (Armouti-Hansen and Kops 2018).

Manzini and Mariotti (2007) assume that a DM forms a sequence of shortlists of decreasing size using a set of criteria ordered from most to least important until an offer is selected. Au and Kawai (2011) present a model of selecting a shortlist on the basis of initial information. More information is then gained about the offers on the shortlist, before a final selection is made. These papers describe general properties of such decision rules. A desirable criterion is given by the Weak Axiom of Revealed Preferences (WARP, see Lleras et al. 2017). This criterion is fulfilled when the following condition is satisfied: if offers x and y belong to both of the sets S and T and x is chosen from S, then y is never chosen from T. In mathematical terms, when $c(S)$ is the offer chosen from S under a given search procedure

$$[c(S) = x, y \in S, x \in T] \Rightarrow [c(T) \neq y]. \tag{1}$$

Search procedures that construct a shortlist of fixed size using initial information do not satisfy WARP, which can be seen as follows. Suppose that offer y is more attractive than offer x based on initial information, but close inspection shows that x is more attractive than y. One can define sets of offers S and T, such that both x and y are on the shortlist of offers from S and x is ultimately chosen, but x is not on the shortlist of offers from set T and y is chosen. Although such procedures do not satisfy WARP, when close inspection of offers is costly, they can significantly reduce search costs compared with procedures satisfying WARP.

Few articles present optimization models involving shortlists. Ramsey (2019) presents a model under which a DM forms a shortlist based on initial information. The offers on the shortlist are then observed closely and the highest ranked offer on the shortlist, according to all the information obtained, is accepted. The DM cannot precisely assess offers, but is able to define a linear ranking of offers using the information available. The DM's strategy determines the number of offers on the shortlist, n. The search costs are convex in n. Under quite weak assumptions about the information regarding the value of an offer, the marginal gain from increasing the length of the shortlist from n to $n + 1$ (the rise in the expected value of the offer chosen) is decreasing in n. Hence, the optimal length of the shortlist is the minimum value of n satisfying the following condition: the marginal gain from increasing n is not greater than the associated marginal rise in search costs. Olkin and Stephens (1993) model the choice of a shortlist of candidate employees when the goal is to hire the best candidate.

The approach presented here addresses some of the shortcomings in the model of Ramsey (2019). Above all, that model does not address the fact that selections are made on the basis of several traits. Secondly, the role of a shortlist in investigating a variety of offers is not considered. Various multicriteria approaches to assessing offers can be applied. We compare two such approaches: SAW (Simple Additive Weighting) and TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution). According to SAW (see Keeney et al. 1993, Sahir et al. 2017), each offer is assessed using a weighted sum of scores, where each score is based on a linear or triangular function of a single trait. TOPSIS (see Yoon and Hwang 1995) assesses an offer based on its distances from the ideal offer and the “anti-ideal” offer in a space defined by the traits considered. Górecka et al. (2016) define the MARS (Measuring Attractiveness around Reference Solutions) approach, which may be interpreted as a generalization of TOPSIS. Decisions based on more complex approaches can be interpreted as choices made using the SAW method according to appropriately defined weights (see Kaliszewski and Podkopaev 2016). The importance (weights) of the traits can be defined in various manners. Here, we consider two approaches: a) the weights are specifically set by the DM, b) the weights are defined on the basis of verbal comparisons of pairs of traits using the Analytic Hierarchy Process (AHP, see Saaty and Peniwati 2008). Mariański et al. (2020) consider the construction of shortlists based on assessments of offers using the SAW method.

As databases of offers can be large, automatic selection of shortlists may be very useful to DMs. Since databases often contain qualitative information about offers (e.g. photos), such a shortlist may also be used as an initial step in deciding which offers to physically view. The traits used to assess offers are classified into three categories: incentives (larger values are more attractive), disincentives (larger values are less attractive) and goal traits, (a target value is preferred, but specified variation is permissible). The minimum and maximum acceptable values of goal traits can be easily defined. In addition, the minimum (maximum) acceptable value for each incentive (disincentive, as appropriate) must be given. The DM should define the relative weights (importance) of each trait or, at least, verbally compare the importance of each pair of traits. Finally, the length of the shortlist should be chosen.

Section 1 presents the general framework. The implementation of the framework for selecting a shortlist is illustrated in Section 2 on the basis of a simple example. Section 3 compares the results obtained using such a procedure to select a shortlist of real estate offers. Finally, some conclusions and directions for future research are given.

1 General Framework

The algorithm constructs shortlists using multivariate data that include continuous variables. The length of the shortlist is chosen by the DM. The DM gives a set of hard constraints that must be satisfied. Such constraints may also relate to discrete and qualitative variables. The initial filtering of offers based on hard constraints is illustrated in Section 2. Continuous traits are classified as incentives, disincentives or goal traits. The offers on the shortlist should satisfy the following two criteria: a) each offer is potentially attractive based on the available information, b) they should exhibit diversity. The second condition may enable a DM to make a more informed final choice, especially when he/she possesses limited information about the environment of the market, e.g., if the DM is looking for a flat in a new city. Thus the algorithm aims to maximize a weighted average of two measures: a) the attractiveness of offers on the shortlist and b) their diversity, which are defined in the following subsections.

2022 Volume XXII(1): 109-120
DOI: 10.25142/aak.2022.009

1.1 Definition of the attractiveness measure

Assume that *I* offers satisfy the hard constraints and each is assessed using *J* continuous traits. Let $x_{i,j}$ be the value of the *j*-th trait for the *i*-th offer. Each observation is transformed to a value $r_{i,j}$ in the interval [0,1] according to the type of trait. If the *j*-th trait is an incentive, then

$$r_{i,j} = \frac{x_{i,j} - x_{min,j}}{x_{max,j} - x_{min,j}}, \tag{1}$$

where $x_{min,j}$ and $x_{max,j}$ are the minimum and maximum values, respectively, of the *j*-th trait (among the offers satisfying the hard constraints). If the *j*-th trait is a disincentive, then

$$r_{i,j} = \frac{x_{max,j} - x_{i,j}}{x_{max,j} - x_{min,j}}. \tag{2}$$

If the *j*-th trait is a goal trait where $x_{pref,j}$ is the preferred value, let $x_{maxd,j}$ be the maximum deviation from the preferred value (among the offers satisfying the hard constraints). Then,

$$r_{i,j} = \frac{x_{maxd,j} - |x_{i,j} - x_{pref,j}|}{x_{maxd,j}}. \tag{3}$$

Let w_j be the weight of the *j*-th trait. These weights can be set by the DM or determined by comparing the importance of each pair of traits. Suppose that trait *j* is stated to be at least as important as trait *k*. Then the score ascribed to trait *j*, $s_{j,k}$, is defined according to Table 1. The score given to trait *k* is $s_{k,j} = 1/s_{j,k}$. By definition $s_{j,j} = 1$. Even scores can be interpreted as the relative importance of traits lying between the two appropriate verbal descriptions. Two methods of defining weights using such scores are presented in Section 2.

Table 1: Comparison of the importance of traits

Score: $s_{j,k}$	Interpretation
1	Traits *j* and *k* are equally important
3	Trait *j* is slightly more important that trait *k*
5	Trait *j* is more important than trait *k*
7	Trait *j* is strongly more important than trait *k*
9	Trait *j* is absolutely more important than trait *k*

Source: Chiara Mocenni (via https://www3.diism.unisi.it/~mocenni/Note_AHP.pdf) accessed 23/07/2021

The value $r_{i,j}$ is a standardized measure of offer *i*'s attractiveness according to trait *j*. Using SAW, the attractiveness of offer *i* based on the initial information is given by A_i^S, where

$$A_i^S = \sum_{j=1}^{J} w_j r_{i,j} \tag{4}$$

Using TOPSIS, the final value of $r_{i,j}$ is calculated by multiplying the scaled variable, defined above, by w_j. Hence, $r_{i,j} = w_j$ may be interpreted as the "ideal" value of a trait and $r_{i,j} = 0$ as the "anti-ideal" value. Thus the vector $(w_1, w_2, \dots, w_J)$ corresponds to an "ideal" offer, while $(0,0,\dots,0)$ corresponds to an "anti-ideal" offer. We may calculate the distances between the *i*-th offer and both the anti-ideal offer, S_i^+, and the ideal offer, S_i^-, as follows:

$$S_i^+ = \sqrt{r_{i,1}^2 + r_{i,2}^2 + \dots + r_{i,J}^2}\ ; \ \ S_i^- = \sqrt{(w_1 - r_{i,1})^2 + (w_2 - r_{i,2})^2 + \dots + (w_J - r_{i,J})^2}\,. \tag{5}$$

An attractive offer is similar to the ideal, while being dissimilar to the anti-ideal. Hence, we define the attractiveness of the i-th offer, A_i^T, by

$$A_i^T = \frac{S_i^+}{S_i^+ + S_i^-}. \tag{6}$$

1.2 Definition of the diversity measure

Suppose that the diversity of a set of offers, defined as a standardized mean distance between pairs of offers, is assessed on the basis of K continuous traits. These could be the same traits as those used to measure attractiveness. However, suppose that one trait used to assess an offer is the distance of a flat from the DM's workplace. The diversity of offers should be based on the physical distance between pairs of flats and not on the "differences between distances to the workplace" (two flats located at the same distance from the workplace might be in completely different districts). In general, it suffices that the sets of traits defining both the attractiveness of an offer and the diversity of offers can be obtained from the database.

Let $y_{i,k}$ be the observation for the i-th offer of the k-th trait used in defining the distance between offers. Each of these traits is scaled to the interval [0, 1] as in Equation (1), i.e. set

$$s_{i,k} = \frac{y_{i,k} - y_{min,k}}{y_{max,k} - y_{min,k}}. \tag{7}$$

The distance between the i-th and m-th offers, $d_{i,m}$, can initially be defined as the Euclidean distance between the vectors $\boldsymbol{s}_i$ and $\boldsymbol{s}_m$, where $\boldsymbol{s}_i = (s_{i,1}, s_{i,2}, \dots, s_{i,K})$. Hence,

$$d_{i,m} = \sqrt{(s_{i,1} - s_{m,1})^2 + (s_{i,2} - s_{m,2})^2 + \cdots + (s_{i,K} - s_{m,K})^2}\,. \tag{8}$$

These distances are not standardized (the maximum possible distance is $\sqrt{K}$). Dividing each distance by the largest distance between offers gives a standardized value.

1.3 Construction of a shortlist

Let the weight of the diversity of a shortlist be α. Define the score of the shortlist of offers $N, v(N)$, to be $v(N) = (1 - \alpha)\bar{x}_N + \alpha\bar{d}_N$, where $\bar{x}_N$ is the mean attractiveness of the offers in N and $\bar{d}_N$ is the mean of the standardized distances between pairs of offers in N. A shortlist of n offers is constructed by sequentially adding offers using the following greedy algorithm:

1. Let $c = 1$. The offer with the highest attractiveness score is placed on the list.
2. Add the offer not yet placed on the list that maximizes the score of the new list.
3. Let $c = c + 1$. If $c = n$, then stop, otherwise return to 2.

If no weight is ascribed to diversity, then the n most attractive offers are placed on the shortlist. In this case, the algorithm maximizes the score of the shortlist (the mean of the attractiveness measures). However, if $\alpha > 0$, the shortlist formed is not necessarily optimal, since its score cannot be defined as the "sum of the individual gains made at each stage".

2 A Simple Example

Suppose that a DM is looking for a flat. The following information on offers is available from an Internet site: a) price, b) size (in m^2), c) no. of bedrooms, d) latitude relative to the DM's

2022 Volume XXII(1): 109-120
DOI: 10.25142/aak.2022.009

workplace ("distance north" - in kms), e) longitude relative to the DM's workplace ("distance east" - in kms). The DM's hard constraints and the types of the corresponding variables are:

- **Price:** below 500 000 (disincentive).
- **Size:** preferred size 80m^2, $\pm$ 20m^2 (target).
- **No. of bedrooms**: between 2 and 4, inclusively (only used for initial filtering).
- **Distance from workplace**: less than 10km (disincentive).

Data about eight offers are given in Table 2. The distance from the workplace is defined as the Euclidean distance given by two positional coordinates. The DM wishes to select three flats to view. The number of bedrooms is only used in filtering based on the hard constraints.

Table 2: Data regarding flats for sale

	Price	Size	No. of bedrooms	Distance N	Distance E
Flat 1	450 000	84	4	2.5	-1.4
Flat 2	390 000	68	3	1.4	-0.9
Flat 3	440 000	63	2	-0.8	0.6
Flat 4	480 000	108	5	2.1	3.2
Flat 5	420 000	76	3	3.1	1.4
Flat 6	450 000	112	4	6.8	-9.3
Flat 7	410 000	88	3	4.0	3.6
Flat 8	490 000	72	3	-1.0	1.2

Source: Example data given by the authors

Based on the hard constraints, Flat 4 and Flat 6 are removed (e.g. they are too large).

As noted in Section 1, the DM can define appropriate weights for the traits, either directly or indirectly using pairwise comparisons. Suppose that pairwise comparisons give the scores $s_{1,2} = 4, s_{1,3} = 8, s_{2,3} = 5$. The resulting matrix of scores is *A*, where

$$A = \begin{pmatrix} 1 & 4 & 8 \\ 0.25 & 1 & 5 \\ 0.125 & 0.2 & 1 \end{pmatrix}. \tag{9}$$

Note that row *i* corresponds to the scores ascribed to trait *i*.

One can define the weights of traits by normalizing the scores. The entries in each column are divided by the sum of the entries in that column. The weight of trait *i* is the mean of the normalized scores in row *i*. Hence, $w_1 \approx 0.6893, w_2 \approx 0.2438, w_3 \approx 0.0669$.

One can also define the weights according to the eigenvector associated with the largest eigenvalue of the matrix of scores. The largest eigenvector is 3.0940 and the eigenvector corresponding to this eigenvector (standardized so that the sum of the components is one) gives $w_1 \approx 0.6986, w_2 \approx 0.2370, w_3 \approx 0.0643$. When the comparisons are relatively consistent (see Kułakowski, 2015), these two methods provide weights of very similar values.

The scaled data used to assess these offers are presented in Table 3. For the purposes of illustration, these traits are equally weighted (i.e. $w_i = 1/3, i = 1, 2, 3$). The data used to calculate the diversity of offers are given in Table 4. The matrix of distances and standardized distances between offers is presented in Table 5. We construct a shortlist of three flats when the weights ascribed to attractiveness and diversity are 10/13 and 3/13, respectively. TOPSIS is applied to assess the offers. The procedure based on SAW is analogous. First, the highest ranked offer (Flat 2) is placed on the shortlist. Next, we add a second offer to maximize the score of

2022 Volume XXII(1): 109-120
DOI: 10.25142/aak.2022.009

the shortlist. This is illustrated in Table 6. On the basis of these calculations, Flat 7 is added to the shortlist. Finally, a third flat is added to the shortlist. This is illustrated in Table 7. It follows that the final shortlist consists of Flats 2, 7 and 8.

The number of possible shortlists of three flats is ${}_6C_3 = \frac{6!}{3!3!} = 20$. Thus exhaustive search could be easily used to find the optimal shortlist. However, when many offers satisfy the hard constraints compared to the length of the shortlist, the possible number of shortlists is approximately $I^n/n!$. In comparison, the number of shortlists considered by the greedy algorithm is much smaller (of order In). Exhaustive search indicates that, in this case, the greedy algorithm finds the optimal shortlist of three offers based on the weighted score.

Table 3: Data used to assess the attractiveness of flats (scaled data given in brackets)

	Price	Size	Distance from workplace	A_i^S	S_i^+	S_i^-	A_i^T
Flat 1	450 000 (0.4)	84 (1)	2.8653 (0.5743)	0.6581	0.4069	0.2452	0.6239
Flat 2	390 000 (1)	68 (0.3846)	1.6643 (0.8484)	0.7443	0.4555	0.2113	0.6832
Flat 3	440 000 (0.5)	63 (0)	1.0000 (1)	0.5000	0.3727	0.3727	0.5000
Flat 5	420 000 (0.7)	76 (1)	3.4015 (0.4519)	0.7173	0.4339	0.2083	0.6757
Flat 7	410 000 (0.8)	88 (0.6923)	5.3814 (0)	0.4974	0.3527	0.3551	0.4983
Flat 8	490 000 (0)	72 (0.6923)	1.5620 (0.8717)	0.5213	0.3711	0.3514	0.5136

Source: Example data given by the authors (based on Table 2)

Table 4: Data used to assess the diversity of offers (scaled data given in brackets)

	Price	Size	Distance N	Distance E
Flat 1	450 000 (0.6)	84 (0.84)	2.5 (0.7)	-1.4 (0)
Flat 2	390 000 (0)	68 (0.2)	1.4 (0.48)	-0.9 (0.10)
Flat 3	440 000 (0.5)	63 (0)	-0.8 (0.04)	0.6 (0.4)
Flat 5	420 000 (0.3)	76 (0.52)	3.1 (0.82)	1.4 (0.56)
Flat 7	410 000 (0.2)	88 (1)	4.0 (1)	3.6 (1)
Flat 8	490 000 (1)	72 (0.36)	-1.0 (0)	1.2 (0.52)

Source: Example data given by the authors (based on Table 2)

Table 5: Matrix of distances between offers

	Flat 2	Flat 3	Flat 5	Flat 7	Flat 8
Flat 1	0.9099 (0.5909)	1.1451 (0.7436)	0.7214 (0.4684)	1.1294 (0.7334)	1.0728 (0.6966)
Flat 2		0.7574 (0.4918)	0.7208 (0.4681)	1.3268 (0.8616)	1.1968 (0.7772)
Flat 3			0.9718 (0.6310)	1.5400 (1.0000)	0.6290 (0.4084)
Flat 5				0.6829 (0.4435)	1.0907 (0.7082)
Flat 7					1.5100 (0.9805)

Source: Example data given by the authors (based on Table 2)

Table 6: Step two of constructing a shortlist

Shortlist	Attractiveness scores	Mean attractiveness	Standardized Distance between offers	Weighted Score
{2, 1}	{0.6832, 0.6239}	0.65355	0.5909	0.6391
{2, 3}	{0.6832, 0.5000}	0.59160	0.4918	0.5686
{2, 5}	{0.6832, 0.6757}	0.67945	0.4681	0.6307
{2, 7}	**{0.6832, 0.4983}**	**0.59075**	**0.8616**	**0.6533**
{2, 8}	{0.6832, 0.5136}	0.59840	0.7772	0.6397

Source: Example data given by the authors (based on Table 2)

Table 7: Step three of constructing a shortlist

Shortlist	Attractiveness scores	Distance between offers*	Weighted score
{2, 7, 1}	{0.6832, 0.4983, 0.6239}	{0.8616, 0.5909, 0.7334}	0.6311
{2, 7, 3}	{0.6832, 0.4983, 0.5000}	{0.8616, 0.4918, 1.0000}	0.6122
{2, 7, 5}	{0.6832, 0.4983, 0.6757}	{0.8616, 0.4681, 0.4435}	0.6126
{2, 7, 8}	**{0.6832, 0.4983, 0.5136}**	**{0.8616, 0.7772, 0.9805}**	**0.6361**

Source: Example data given by the authors (based on Table 2). * - The three values are the distances between a) the first and second in the list, b) the first and third on the list and c) the second and the third on the list, respectively

Applying the algorithm in conjunction with SAW, the shortlist chosen is {1,2,7}. Its score is 0.6553. Exhaustive search indicates that the optimal shortlist and score are {1,2,5} and 0.6610,

2022 Volume XXII(1): 109-120
DOI: 10.25142/aak.2022.009

respectively. The shortlist selected by the greedy algorithm is the third best according to the given criterion. From Table 3, it can be seen that the offers are ascribed identical ranks by both methods, while the range of attractiveness scores obtained using SAW is greater. Since the diversity measure does not depend on the method used to assess the offers, it is intuitive that the optimal shortlist based on SAW contains offers that are generally better ranked than the list obtained using TOPSIS. Flats 1, 2 and 7 (chosen using SAW) are ranked 1, 3 and 6, respectively. Using this criterion, the optimal shortlist actually includes the top three ranked offers. On the other hand, Flats 2, 7 and 8 (chosen using TOPSIS) are ranked 1, 4 and 6, respectively (this is the optimal shortlist in this case). However, this phenomenon might result from the specificity of the dataset. The choice of both the method of assessment and the weight ascribed to diversity will be investigated more thoroughly in future research.

3 A Practical Example

On 4th Feb. 2021 data regarding 11 321 offers of properties in the city of Wrocław were downloaded from the webpage www.otodom.pl. The data used were: price, size (in m^2) and number of rooms. The goal of the analysis is to propose a shortlist of 30 offers using TOPSIS and SAW. The hard constraints that an offer must satisfy are as follows:

1. **Price:** cannot exceed 500 000PLN (approx. €111 500). This is a disincentive.
2. **Size:** target 80m^2, permissible variation $\pm 20\text{m}^2$.
3. **No. of rooms:** at least three (only used for initial filtering).

After eliminating the offers that did not satisfy these constraints, 1 176 offers remained. The weights ascribed to price and size were 0.3 and 0.7, respectively. The algorithm described in Section 1 was applied in conjunction with SAW and TOPSIS separately to derive shortlists of 30 flats. Table 8 presents the mean rank, attractiveness and diversity of the offers selected according to both the method of assessment and weight ascribed to diversity, where $\alpha \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$.

The mean measure of attractiveness varies according to the method applied. Hence, relative measures of the mean attractiveness of offers on the shortlist and their diversity are used to enable comparison. These relative measures are defined as the ratio between the given mean and the maximum of that mean (for SAW and TOPSIS individually). The maximum mean attractiveness occurs when $\alpha = 0$, while the maximum diversity occurs when $\alpha = 1$. The position "no. of offers in common" gives the size of the set of offers that are common to the shortlists derived using the two methods of assessment for a fixed value of α.

When $\alpha = 0$, the 30 highest-ranked offers are placed on the shortlist. All but two (the 29th and 30th) of the 30 offers ranked highest by TOPSIS are ranked in the top thirty using SAW. The two offers ranked in the top 30 by SAW, but not by TOPSIS, are ranked 27th and 30th. Hence, the rankings based on these methods are very highly correlated. When $\alpha = 1$, the algorithm first selects the highest-ranked offer and then successively adds the offer which maximizes diversity. Since the diversity measure is independent of the method of assessment, whenever the assessment methods rank the same offer most highly, the shortlists formed are identical. In the problem considered, the offer ranked first by TOPSIS is ranked second by SAW (and vice versa). The only difference between these two shortlists lies in the first offer chosen (the highest ranked by the appropriate method). The offer ranked highest by TOPSIS is a 78m^2 flat costing 425 000PLN, while the offer ranked highest by SAW is a 79m^2 flat costing 443 980PLN. As these offers are very similar in both dimensions, it is unsurprising that the remaining offers placed on the list are independent of the starting point. It seems theoretically possible that

2022 Volume XXII(1): 109-120
DOI: 10.25142/aak.2022.009

shortlists constructed in this way could be completely different. However, this is a theoretical consideration, since maximizing the diversity of offers without considering their attractiveness (apart from including the highest ranked offer) seems of little practical use.

Table 8: The effect of the weight placed on diversity to the mean ranks, attractiveness and diversity of real estate offers placed on a shortlist of length 30 (1 176 offers satisfy the hard constraints). * It should be noted that the difference between the mean ranks obtained when $\alpha = 0$ results from offers having equal measures of attractiveness according to TOPSIS.

		TOPSIS	SAW
$\alpha = 0$	Mean rank (Max. rank)	15.7000 (32.5)	15.5000 (30)
	Mean attractiveness (mean diversity)	0.7121 (0.1342)	0.6933 (0.1424)
	Relative attractiveness (relative diversity)	1.0000 (0.2585)	1.0000 (0.2744)
	No. of offers in common	28	-
$\alpha = 0.1$	Mean rank (Max. rank)	16.5000 (49)	17.0667 (60)
	Mean attractiveness (mean diversity)	0.7115 (0.1474)	0.6913 (0.1698)
	Relative attractiveness (relative diversity)	0.9991 (0.2839)	0.9971 (0.3273)
	No. of offers in common	28	-
$\alpha = 0.2$	Mean rank (Max. rank)	21.4167 (113)	24.4333 (138)
	Mean attractiveness (mean diversity)	0.7046 (0.1841)	0.6795 (0.2302)
	Relative attractiveness (relative diversity)	0.9893 (0.3547)	0.9802 (0.4437)
	No. of offers in common	27	-
$\alpha = 0.3$	Mean rank (Max. rank)	39.0000 (338)	33.6667 (301)
	Mean attractiveness (mean diversity)	0.6819 (0.2560)	0.6687 (0.2674)
	Relative attractiveness (relative diversity)	0.9576 (0.4933)	0.9646 (0.5155)
	No. of offers in common	29	-
$\alpha = 0.4$	Mean rank (Max. rank)	78.2000 (465)	65.1667 (416)
	Mean attractiveness (mean diversity)	0.6383 (0.3387)	0.6307 (0.3403)
	Relative attractiveness (relative diversity)	0.8962 (0.6527)	0.9098 (0.6558)
	No. of offers in common	26	-
$\alpha = 0.5$	Mean rank (Max. rank)	142.6833 (540)	121.1000 (560.5)
	Mean attractiveness (mean diversity)	0.5756 (0.4153)	0.5746 (0.4061)
	Relative attractiveness (relative diversity)	0.8083 (0.8002)	0.8288 (0.7827)
	No. of offers in common	25	-
$\alpha = 0.6$	Mean rank (Max. rank)	204.2333 (1176)	188.7333 (1176)
	Mean attractiveness (mean diversity)	0.5224 (0.4581)	0.5203 (0.4496)
	Relative attractiveness (relative diversity)	0.7336 (0.8827)	0.7506 (0.8865)
	No. of offers in common	26	-
$\alpha = 0.7$	Mean rank (Max. rank)	336.1667 (1176)	325.9000 (1176)
	Mean attractiveness (mean diversity)	0.4402 (0.5024)	0.4317 (0.4985)
	Relative attractiveness (relative diversity)	0.6181 (0.9681)	0.6228 (0.9607)
	No. of offers in common	26	-
$\alpha = 0.8$	Mean rank (Max. rank)	386.9333 (1176)	368.0333 (1176)
	Mean attractiveness (mean diversity)	0.4107 (0.5118)	0.4024 (0.5092)
	Relative attractiveness (relative diversity)	0.5767 (0.9862)	0.5805 (0.9815)
	No. of offers in common	28	-
$\alpha = 0.9$	Mean rank (Max. rank)	460.3667 (1176)	424.3667 (1176)
	Mean attractiveness (mean diversity)	0.3704 (0.5184)	0.3703 (0.5160)
	Relative attractiveness (relative diversity)	0.5201 (0.9989)	0.5341 (0.9945)
	No. of offers in common	28	-
$\alpha = 1.0$	Mean rank (Max. rank)	498.2667 (1176)	499.4333 (1176)
	Mean attractiveness (mean diversity)	0.3490 (0.5190)	0.3300 (0.5188)
	Relative attractiveness (relative diversity)	0.4900 (1.0000)	0.4759 (1.0000)
	No. of offers in common	29	-

Source: A program written in R by the authors to construct shortlists of attractive offers using data from the Internet site www.otodom.pl (accessed Feb. 4th, 2021)

Optimizing the weight ascribed to diversity is of major practical importance. For $\alpha \leq 0.3$, the relative attractiveness of the offers on the shortlist is greater than 0.95. Closer investigation

2022 Volume XXII(1): 109-120
DOI: 10.25142/aak.2022.009

indicates that the twenty highest ranked offers are always selected for such values of α. Additionally, when α is between 0.1 and 0.3, several offers selected significantly differ from the highest ranked offers, but are relatively well ranked (within the top 30% of the offers satisfying the hard constraints). Hence, using such values of α may be appropriate. It is interesting to compare the diversity of the offers satisfying the hard constraints with the diversity of the offers placed on a shortlist. The diversity of the offers that satisfy the hard constraints is 0.2277. Using TOPSIS to assess the attractiveness of offers, the diversity of offers on the shortlist for $\alpha = 0.1, 0.2, 0.3$ is 0.1474, 0.1841 and 0.2560, respectively. Using SAW, the analogous diversity measures are 0.1698, 0.2302 and 0.2674. The program takes about 5 seconds to run for a given value of α on a laptop with a 1.4GH Intel Core i5 processor. Hence, one can propose the following rule of thumb for choosing a shortlist:

1. Construct shortlists using about 10 relatively low values of α.
2. Select the shortlist for which the mean distance between offers is most similar to the mean distance between offers that satisfy the hard constraints.

This is a heuristic procedure for maximizing the mean attractiveness of offers on the shortlist while preserving their diversity. This approach seems logical when the DM has limited knowledge of a market, otherwise a slightly smaller value of α might be appropriate. Additionally, appropriate values of the parameter α may depend on the preferences of the DM and the joint distribution of the traits used. This will be investigated in future research.

Conclusion

We have presented an original algorithm that constructs shortlists of attractive offers from large databases. This approach is very practical when a DM wishes to purchase a unique good for which some information can be accessed easily using a database, but offers need to be carefully inspected before purchase. The size of this shortlist, together with hard constraints that an offer must satisfy, are defined by the DM. Firstly, the offers that do not satisfy the hard constraints are removed. The remaining offers are then evaluated using a method for multicriteria assessment. Finally, a shortlist is constructed by an algorithm, whose goal is to maximize a weighted average of the mean assessments of the offers on the shortlist and their diversity. Such a procedure is practical in deciding which offers to physically observe.

This algorithm was illustrated using an example in which a shortlist of flats is chosen. The results show that both SAW and TOPSIS can be successfully adapted to select such shortlists. For these data, when a weight of between 0.1 and 0.3 is ascribed to the diversity of offers on the shortlist, this algorithm produces shortlists which include highly ranked offers that exhibit a similar level of diversity to those satisfying the hard constraints. Since a suitable weighting procedure may depend on the preferences of the DM and the joint distribution of the traits considered, a heuristic procedure for selecting the value of the weight, α, is suggested. Essentially, this procedure aims to maximize the mean attractiveness of the offers on the shortlist, while preserving the diversity of the offers that satisfy the hard constraints. The robustness of such a procedure will be investigated in future research.

The set of variables used may be expanded relatively easily. Since the measures of attractiveness and diversity are scaled to the range [0, 1], this method should be robust to the number of variables used. Also, the relationship between the time required by the algorithm and the number of variables used only depends on checking whether the hard constraints are satisfied and calculating the attractiveness and distance measures. The complexity of these procedures is linear in the number of variables used. Hence, the inclusion of extra variables should not be problematic. This will be checked in a future extensive study.

Various multicriteria methods for assessing the attractiveness of an offer are available. We have considered the use of TOPSIS and SAW to construct a shortlist of real estate offers. The shortlists formed using these two approaches were very similar. Since SAW is easier to understand/interpret, it might be more practical to implement an algorithm based on this approach. In the future, we will compare procedures for selecting shortlists over a wider range of problems, in order to propose a robust approach to such problems.

Acknowledgement
David Ramsey and Aleksander Mariański are grateful to funds obtained from the Polish National Science Centre for project number 2018/29/B/HS4/02857, "Logistics, Trade and Consumer Decisions in the Age of the Internet".

This article is an expanded version of work presented at the DESME 2021 (Decision Making for Small and Medium-Sized Enterprises) meeting, May 13th-14th, Prague, Czech Republic.

References

[1] ARMOUTI-HANSEN, J. and C. KOPS, 2018. This or That? Sequential Rationalization of Indecisive Choice Behavior. *Theory and Decision*, **84**(4), 507–524. ISSN 0040-5833.

[2] AU, P. H. and K. KAWAI, 2011. Sequentially Rationalizable Choice with Transitive Rationales. *Games and Economic Behavior*, **73**(2), 608–614. ISSN 0899-8256.

[3] BELTON, V., 1985. The Use of a Simple Multiple-Criteria Model to Assist in Selection from a Shortlist. *Journal of the Operational Research Society*, **36**(4), 265–274. ISSN 0160-5682.

[4] BORAH, A. and C. KOPS, 2019. Rational Choices: An Ecological Approach. *Theory and Decision*, **86**(3–4), 401–420. ISSN 0040-5833.

[5] DURBACH, I. and S. DAVIS, 2012. Decision Support for Selecting a Shortlist of Electricity-Saving Options: A Modified SMAA Approach. *ORiON*, **28**, 2, 99–116. ISSN 0259-191X.

[6] DULLECK, U., F. HACKL, B. WEISS and R. WINTER-EBMER, 2011. Buying Online: An Analysis of Shopbot Visitors. *German Economic Review*, **12**(4), 395–408. ISSN 1465-6485.

[7] GÓRECKA, D., E. ROSZKOWSKA and T. WACHOWICZ, 2016. The MARS Approach in the Verbal and Holistic Evaluation of the Negotiation Template. *Group Decision and Negotiation*, **25**(6), 1097–1136. ISSN 0926-2644.

[8] KALISZEWSKI, I. and D. PODKOPAEV, 2016. Simple Additive Weighting—A Metamodel for Multiple Criteria Decision Analysis Methods. *Expert Systems with Applications*, **54**, 155–161. ISSN 0957-4174.

[9] KEENEY, R. L, H. RAIFFA and R. F. MEYER, 1993. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge, UK: Cambridge University Press. ISBN 978-0-5214388-3-4.

[10] KUŁAKOWSKI, K., 2015. Notes on Order Preservation and Consistency in AHP. *European Journal of Operational Research*, **245**(1), 333-337. ISSN 0377-2217.

[11] LLERAS, J. S., Y. MASATLIOGLU, D. NAKAJIMA and E. Y. OZBAY, 2017. When More is Less: Limited Consideration. *Journal of Economic Theory*, **170**, 70–85. ISSN 0022-0531.

[12] MANZINI, P. and M. MARIOTTI, 2007. Sequentially Rationalizable Choice. *American Economic Review*, **97**(5), 1824–1839. ISSN 0002-8282.

[13] MARIAŃSKI, A., M. KĘDZIORA, D. M. RAMSEY and L. SZCZUROWSKI, 2020. On forming shortlists of attractive offers from large databases: the example of purchasing a flat. In: K. S. Soliman, ed. *Proceedings of the 35th International Business Information Management Association Conference (IBIMA)*. Seville, Spain: International Business Information Management Association, pp. 13037–13047. ISBN 978-0-9998551-4-1.

[14] MASATLIOGLU, Y., D. NAKAJIMA and E. Y. OZBAY, 2012. Revealed Attention. *American Economic Review*, **102**(5), 2183–2205. ISSN 0002-8282.

[15] OLKIN, I. and M. A. STEPHENS, 1993. On Making the Shortlist for the Selection of Candidates. *International Statistical Review/Revue Internationale de Statistique*, **61**(3), 477–486. ISSN 1751-5823.

[16] RAMSEY, D. M., 2019. Optimal Selection from a Set of Offers Using a Shortlist. *Multiple Criteria Decision Making*, **14**, 75–92. ISSN 2084-1531.

[17] SAATY, T. L. and K. PENIWATI, 2008. *Group decision making: drawing out and reconciling differences*. Pittsburgh, Pennsylvania: RWS Publications. ISBN 978-1-888603-08-8.

[18] SAHIR, S. H., R. ROSMAWATI and K. MINAN, 2017. Simple Additive Weighting Method to Determining Employee Salary Increase Rate. *International Journal Scientific Research in Science and Technol*ogy, **3**(8), 42-48. ISSN 2395-6011.

[19] SCHWARTZ, B., 2004. *The paradox of choice: why more is less.* New York: HarperCollins. ISBN 978-0-0624499-2-4.

[20] SIMON, H. A., 1956. Rational Choice and the Structure of the Environment. *Psychological Review*, **63**(2), 129–138. ISSN 0033-295X.

[21] TODD, P. M. and G. GIGERENZER, 2000. Précis of Simple Heuristics that Make us Smart. *Behavioral and Brain Sciences*, **23**(5), 727–741. ISSN 0140-525X.

[22] YOON, K. P. and C. L. HWANG, 1995. *Multiple attribute decision making: an introduction*. (*Series: Quantitative Applications in the Social Sciences, Vol. 104*). Thousand Oaks, CA: Sage Publications. ISBM 978-0-8039548-6-1.