Vol. 34, No. 3 (2024) | DOI: 10.37190/ord240312

OPEN ACCESS

Operations Research and Decisions

www.ord.pwr.edu.pl

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

A synthetic measure for ranking overall satisfaction based on multidimensional ordinal data. The case of airlines on the TripAdvisor website

Ewa Roszkowska[1] Bartłomiej Jefmański[2*] Marta Kusterka-Jefmańska[3]

[1] *Faculty of Computer Science, Bialystok University of Technology, Białystok, Poland*
[2] *Department of Econometrics and Computer Science, Wroclaw University of Economics and Business, Wrocław, Poland*
[3] *Department of Quality and Environmental Management, Wroclaw University of Economics and Business, Wrocław, Poland*
[*] *Corresponding author, email address: bartlomiej.jefmanski@ue.wroc.pl*

Abstract

A multi-criteria method has been proposed to measure global consumer satisfaction by handling graphical information represented in circles or star ratings available on the website for several criteria. The Synthetic Measure for Ordinal Data (SMOD) is based on Hellwig's framework measuring distances between the object and the ideal object. In the construction of the synthetic measure, the distance between the ideal and anti-ideal object is also used. GDM2 distance measure was used for the calculation of distances between objects evaluated by ordinal scale. The proposed framework can be an alternative for the fuzzy or intuitionistic fuzzy transformations of linguistic data into fuzzy context. The SMOD method was applied for evaluation and ranking airlines based on passengers' reviews accessed on the TripAdvisor website. We found that the proposed multi-criteria method enables us to discriminate and rank airlines based on incomplete information presented on this website. According to SMOD, the best is Singapore Airlines, while the last is Air China when considering Star Alliance members.

Keywords: *online consumer review, uncertainly, incomplete information, multi-criteria method, ordinal scale, synthetic measure, airlines ranking, TripAdvisor*

1. Introduction

Service quality and consumer satisfaction are important elements of modern business performance and marketing strategy [31, 38]. Several e-commerce websites and social media platforms offer an online information channel that allows customers to post their comments on the products or fill online consumer reviews (OCRs) [15]. Yan et al. [55] pointed out that such opinions may strongly impact customer purchasing decisions, reducing the uncertainty about a product. Electronic word of mouth (e-WOM) [15, 17, 25],

Received 30 July 2023, accepted 24 July 2024, published online 17 October 2024
ISSN 2391-6060 (Online)/

The costs of publishing this issue have been co-financed by the program *Development of Academic Journals* of the Polish Ministry of Education and Science under agreement RCN/SP/0241/2021/1

is realized by e-commerce sites like Amazon and eBay, popular apps like Yelp, Nextdoor, TripAdvisor, Skytrax, and social media platforms like Instagram, Facebook, Twitter, Tik Tok, and YouTube are recognized as an important source of information when making consumer decisions.

Several studies have addressed airline service quality and passenger satisfaction issues. These studies often utilize a combination of quantitative and qualitative research methods. A study by Bellizzi et al. [6] focuses on a literature review of airport and airline service quality, particularly examining methodology and measurement tools. The review encompasses papers published between 2008 and 2018 in journals indexed on databases such as Scopus or Web of Science. Another study by Eboli et al. [14] conducts a review of existing literature on air transport service quality, considering the perspective of passengers. This review includes papers related to airline and airport services published between 2010 and 2020, with the journals being indexed in databases like Scopus and Web of Science. Also, Gupta [18] conducted a literature review focusing on research related to airline service quality. The majority of the studies analyzed were centered around the Service Quality Model (SERVQUAL) [5, 37, 49].

Pakdil and Aydın [37] utilized SERVQUAL scores and factor analysis to assess airline service quality at a Turkish airline. Basfirinci and Mitra [5] employed the SERVQUAL scale and Kano model to measure airline service quality across different cultures. Chu et al. [10] evaluated the quality of service in Taiwanese airlines using the fuzzy-weighted SERVQUAL method.

Certain research studies were oriented toward Structural Equation Modeling (SEM) in the context of airlines. The primary objective of such studies is to analyze the relationships between airline service quality, passenger satisfaction, and other latent constructs. Kos Koklic et al. [28] conducted a survey involving 382 passengers, and their SEM model examined customer satisfaction with both low-cost and full-service airline companies. Suki [46] applied an SEM model to explore the relationship between customer satisfaction with airline service quality and 'word-of-mouth' recommendations, finding them to be consistent. Gursoy et al. [19] investigated the positions of the 10 major US airlines based on 15 attributes measuring actual airline performance concerning quality criteria, applying correspondence analysis.

Several studies have implemented MCDM techniques for assessing airline passenger satisfaction and comparing multiple airlines to establish rankings. A model for airline selection is presented, utilizing the best-worst and VIKOR methodology [18]. The study identifies seven main attributes and twenty-nine sub-attributes related to airline service quality. Tsafarakis et al. [48] recommended the multicriteria satisfaction analysis method (MUSA) for evaluating passengers' satisfaction. The proposed framework is illustrated in the context of Aegean Airlines. Tsaur et al. [49] employ fuzzy set theory to evaluate airline service quality, utilizing the analytic hierarchy process (AHP) for criteria weighting and the technique for order of preference by similarity to ideal solution (TOPSIS) for ranking. The study considers five service quality criteria: tangibility, reliability, responsiveness, assurance, and empathy. Chen [9] focuses on the Taiwanese airline industry, selecting criteria for airline service quality improvement using a model based on decision-making trial and evaluation laboratory (DEMATEL) and analytic network process (ANP) methods. Atalay [1] introduces a hybrid method based on importance-performance-impact analysis (IPIA), fuzzy logic, and information entropy to identify priorities for airline passengers. The study suggests optimal resource allocation strategies to enhance service quality and customer satisfaction. Kuo [29] proposes a framework that combines VIKOR, grey relational analysis (GRA), and interval-valued fuzzy sets to assess the service quality of Chinese cross-strait passenger airlines through customer surveys.

Nowadays, very vital are online consumer reviews about tourism-related products shared on social media platforms such as TripAdvisor, Skytrax. Filieri et al. [16] noted that TripAdvisor's number of monthly visitors has grown from 20 million monthly unique visitors in 2010 to 463 million in 2019. The data obtained from the TripAdvisor website prove to be highly valuable for examining the quality of airline services. In the work of Chang et al. [8] customer satisfaction was analyzed through sentiment analysis and visual analytics, focusing on flight reviews on TripAdvisor spanning from January 2016 to August 2020. Special attention was given to assessing the impact of COVID-19 on passenger travel experiences. Sezgen et al. [44] studied some aspects of customer satisfaction and dissatisfaction concerning full-service and low-cost carriers, as well as economy and premium cabins. They employed latent semantic analysis, a text mining and categorization technique, to analyze online airline reviews on TripAdvisor. Stamolampros et al. [45] utilizing passenger reviews sourced from TripAdvisor, investigated disparities in online ratings given by airline passengers to domestic and foreign carriers.

On websites, consumer feedback is usually provided in terms of rating questions using a Likert scale or textual comments. The powerful format for aggregating individual opinions is circle or star ratings. In this context, there is an urgent need for the development of a tool that allows for accurate, comprehensive analyses of consumer satisfaction using website surveys. The paper aims to present the multi-criteria method to measure global consumer satisfaction by handling graphical information represented in circles or star ratings available on the website for several criteria. Synthetic measure for ordinal data (SMOD) is based on Hellwig's framework measuring distances between the object and the ideal object. The distance between the ideal and anti-ideal object is also used in the construction of this measure. Firstly, the problem of assessing global satisfaction is structured as a multi-criteria problem where data are represented on an ordinal scale. Subsequently, the ideal object is defined as max ratings, while the anti-ideal object as min ratings from evaluated objects. To measure the distance between objects evaluated in the ordinal scale, the concept GDM (generalised distance measure) [50] is applied. Finally, with the principle of the Hellwig method based on two reference points [21] SMOD is calculated which allows for ranking objects.

In the paper, we analyzed the airline service quality performed on passengers' reviews accessed on the TripAdvisor website. The individual passenger's opinions are provided in terms of rating of consumers' experience with the airline in the following aspects: legroom, seat comfort, in-flight entertainment (Wi-Fi, TV, movies), onboard experience, customer service, value for money, cleanliness, check-in and boarding, and food and beverage with the use 5-point scale equivalent to 5 points Likert scale. Next, the aggregated opinions were presented graphically in a circle format and numerically. To rank ordering members of Star Alliance we used SMOD method.

The objectives and contributions of this study can be summarized as follows:

- to develop a modified Hellwig method to handle ordinal data presented in online consumer reviews based on a series of questions,
- to propose the multi-criteria methods that take into consideration unprecise aggregated of consumers' opinions available on the website,

- to present the multi-criteria method which can be an alternative to the fuzzy or intuitionistic fuzzy approach avoiding transformations of linguistic data into a fuzzy context,
- to adopt the distance measure for objects represented by ordinal data for evaluation of consumers' opinions represented by ordinal data.

The rest of the paper is organized as follows. In Section 2, the Star Alliance and source data from the TripAdvisor website used in further analyses are described. The notion of the generalized distance measure (GDM2) and the synthetic measure SMOD, based on ordinal data, are defined. The Section 3 discusses the application of the SMOD method for evaluating customer satisfaction based on online reviews for Star Alliance members. The results of the airline ranking obtained by the SMOD method are compared with the average score of passengers' opinions, the sum of ratings across criteria, and the general rating presented on the TripAdvisor website, demonstrating the usefulness of the proposed approach. The paper concludes with a summary of key findings, limitations and future research.

2. Materials and methods

The following abbreviations have been used in this manuscript:

OCR	–	online consumer review
e-WOM	–	electronic word of mouth
SMOD	–	synthetic measure for ordinal data
SERVQUAL	–	service quality model
MCDM	–	multiple criteria decision-making
SEM	–	structural equation modeling
AHP	–	analytic hierarchy process
ANP	–	analytic network process
TOPSIS	–	technique for order of preference by similarity to ideal solution
VIKOR	–	multi-criteria optimization and compromise solution
	–	viseKriterijumska optimizacija i kompromisno resenje (in Serbian)
MUSA	–	multicriteria satisfaction analysis method
IFSM	–	intuitionistic fuzzy synthetic measure
DIFSM	–	double intuitionistic fuzzy synthetic measure
IPIA	–	importance-performance-impact analysis
I-VIFSM	–	interval-valued intuitionistic fuzzy synthetic measure
GRA	–	grey relational analysis
DEMATEL	–	decision-making trial and evaluation laboratory

2.1. Problem description and data source

Star Alliance was founded in 1997 as the first global aviation alliance. In the beginning, it grouped five airlines: Scandinavian Airlines, Thai Airways International, Air Canada, Lufthansa, and United Airlines offering an international passenger service. Currently, the community includes 26 airlines from around the world with their own distinctive culture and style of service. Star Alliance operates 19,000 flights a day to over 1350 airports in 195 countries.

On TripAdvisor's website, the individual passenger's opinions are provided in terms of rating of consumers' experience with the airline in the following aspects: C1 – legroom, C2 – seat comfort, C3 – in-flight entertainment (Wi-Fi, TV, movies), C4 – onboard experience, C5 – customer service, C6 – value for money, C7 – cleanliness, C8 – check-in and boarding, and C9 – food and beverage. Each of the nine criteria is evaluated graphically by selecting the appropriate number of circles. The circles are additionally described with the following statements: 1 – terrible, 2 – poor, 3 – average, 4 – very good, and 5 – excellent, which are displayed when the mouse hovers over the appropriate circle (see Figure 1). An excerpt from the TripAdvisor website with criteria ratings for Singapore Airlines is shown in Figure 2.

Figure 1. Graphical interface evaluation of airline with respect to the set of criteria from TripAdvisor's website. Source: https://www.tripadvisor.com

The graphical form of evaluating the criteria using circles like in the Figure 2 is easy to read for the consumer but, at the same time, is a source of uncertainty in the measurement process. A large number of reviews requires the calculation of the average ratings for each of the criteria which need the assignment of numerical values to circles. In the case of five circles, these are usually values from 1 to 5. The green colored circle is denoted one point, while half green colored circle 0.5 points (see Figure 2). It should be noted, however, that the assignment of linguistic values to circles may be interpreted differently by consumers, which affects the average rating for selected criteria. Failure to take this fact into account may result, for example, in identical assessments of two facilities in terms of the selected criterion (e.g., at level 4), despite the existence of differences in customer opinions on the quality of this criterion. Therefore, instead of classical multi-criteria methods, their modifications based on fuzzy sets are used, which makes it possible to take into account the uncertainty associated with consumers' perception of linguistic values. The most common approach is to assign triangular or trapezoidal fuzzy numbers to linguistic values [22, 23, 35]. Here, however, there is a problem with the subjectivism of researchers in determining the parameters of fuzzy numbers. Another new approach is to express evaluations of criteria in the form of fuzzy intuitionistic values [26, 27, 30] or interval intuitionistic values [41]. In this case, information on the number of consumer indications of individual categories is required, which in the case of TripAdvisor is the only place for the overall assessment of the airline (criteria ratings are given only as average values).

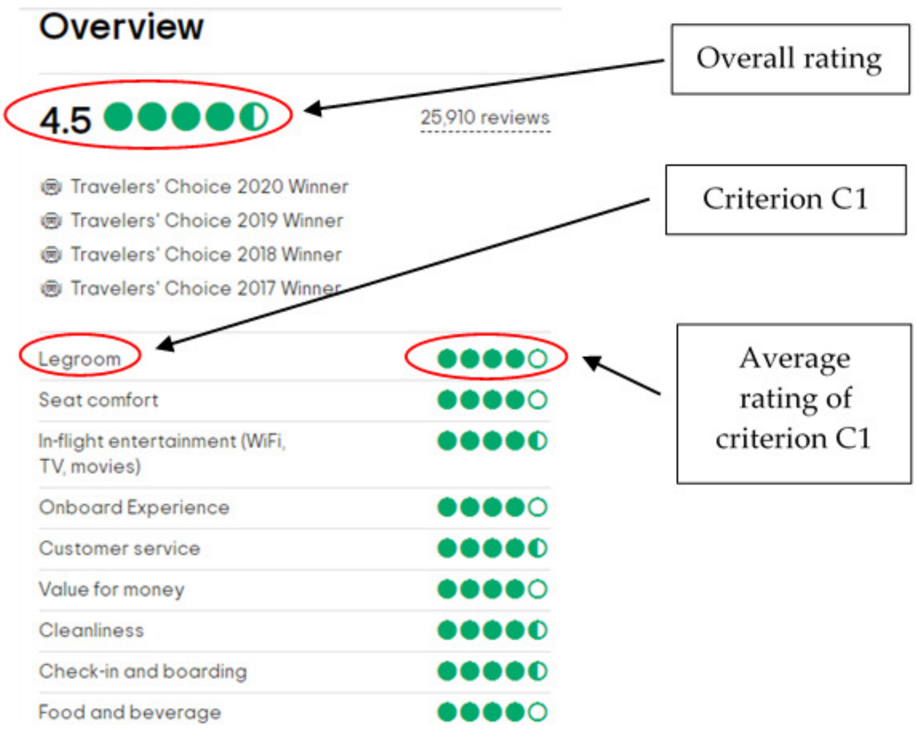

Figure 2. Graphical criteria ratings for Singapore Airlines on TripAdvisor's website. Source: https://www.tripadvisor.com

Therefore, the article proposes another, alternative approach that does not require the transformation of data into a fuzzy form or treating ordinal data as quantitative data (such a scale enhancement is not acceptable from the point of view of measurement theory). The authors' proposal adopts a generalized GDM2 distance measure [50] dedicated to ordinal data in the assessment of the multi-criteria quality of airline services.

2.2. The GDM2 measure for ordinal data. Short overview

Let us note that the distances between any two categories on the ordinal scale are not known, hence for variables measured on the ordinal scale Minkowski distance measure, in particular Euclidean and Hamming distances, cannot be used. Using such distances requires the improper assumption that the distances between adjacent categories on the ordinal scale are equal. Therefore, the definition distance measures should apply acceptable relations on the ordinal scale, i.e., equality, diversity, majority, and minority. Such a measure of distance for ordinal data is the GDM2 proposed by Walesiak [50].

Definition 1. ([50]). Let $OS = \{1, 2, \ldots, k\}$ be a numerical representation of the ordinal scale, where the higher number means more preferable. Let us denote by $O_i = \{O_1, \ldots, O_n\}$ set of objects evaluated on ordinal scale OS with respect m citeria. Let $O_i = [x_{i1}, x_{i2}, \ldots, x_{im}]$, $O_k = [x_{k1}, x_{k2}, \ldots, x_{km}]$ be representation ith and kth object, respectively; $x_{im}, (x_{kj})$ – evaluation ith (kth) object with respect to kth criterion; $x_{ij}, x_{kj} \in OS (j = 1, \ldots, m; i, k = 1, \ldots, n)$. The GDM2 distance between O_i and O_k objects characterized by m criteria has the following form:

$$\mathrm{GDM2}_{ik} = \frac{1}{2} - \frac{\sum_{j=1}^{m} a_{ikj} b_{kij} + \sum_{j=1}^{m} \sum_{l=1;l\neq i,k}^{n} a_{ilj} b_{klj}}{2\left(\left(\sum_{j=1}^{m} a_{ikj}^2 + \sum_{j=1}^{m} \sum_{l=1;l\neq i,k}^{n} a_{ilj}^2\right)\left(\sum_{j=1}^{m} b_{kij}^2 + \sum_{j=1}^{m} \sum_{l=1;l\neq i,k}^{n} b_{klj}^2\right)\right)^{\frac{1}{2}}} \tag{1}$$

and

$$a_{ipj}(b_{krj}) = \begin{cases} 1 & x_{ij} > x_{pj}(x_{kj} > x_{rj}) \\ 0 \quad \text{if} & x_{ij} = x_{pj}, \text{ for } p = k, l; r = i, l \\ -1 & x_{ij} < x_{pj}(x_{kj} < x_{rj}) \end{cases} \tag{2}$$

where $i, k, l = 1, \ldots, n$ – number of objects, $j = 1, \ldots, m$ – number of criteria.

The information on the ordinal scale is incomparably smaller, the only permissible empirical operation on the ordinal scale is the counting of the number of majority, minority, and equality relations. Therefore, in the construction of the distance measure, information about the relations between objects O_i, O_k has been compared to information about the other objects in the set O. In the denominator of the equation (1), the first factor is the number of majority and minority relations specified for the object O_i, and the second factor is the number of majority and minority relations specified for the object O_k.

The distance measure GDM2 has its strengths and weaknesses, the most important of which are:

- can be applied in a situation when variables describing objects are measured only on the ordinal scale,
- needs at least one pair of non-identical objects in the data set not to have zero in the denominator,
- Kendall's idea of correlation coefficient for ordinal variables was used for the GDM2 construction,
- assumes values from the $[0; 1]$ interval. Value 0 indicates that for the compared objects i, k between corresponding observations of ordinal variables, only relations equal to take place. Value 1 indicates that for the compared objects i, k between corresponding observations on ordinal variables, relations "greater than" take place or relations greater than and relations equal to, if they are held for other objects (i.e., objects numbered $l = 1, \ldots, n$ where $l \neq i, k$,
- satisfies conditions: $\mathrm{GDM2}_{ik} \geq 0, GDM2_{ii} = 0, GDM2_{ik} = GDM2_{ki}$ (for all $i, k = 1, \ldots, n$),
- simulation analysis proves that distance does not always satisfy the triangle inequality,
- transformation of ordinal data by any strictly increasing function does not change the value of GDM2 distance.

The generalized distance measure GDM2 was applied for linear ordering objects evaluated in ordinal scale [52]. The paper [51] presents the application of the generalized GDM[1] distance measure for the ranking of universities in Poland. Balcerzak and Pietrzak [4] analyzed the digital economy in Polish regions using TOPSIS method with GDM measure. Dmytrów [13] compared results obtained for several ranking methods, including TOPSIS with GDM measure for the problem of localization selection. Łuczak and Wysocki [34] applied generalized GDM distance measure and TOPSIS methods for assessing the level of socio-economic development of regions of the Wielkopolskie Voivodship.

[1]The GDM designation is used for a variant of the generalized distance measure for the interval or ratio measurement scale.

2.3. Synthetic measure for ordinal data

The taxonomic measure of development (TMD) has been introduced by Hellwig in 1968 [20]. TMD allows ranking objects from the worst to the best in the multidimensional space taking into consideration several criteria. This method is based on the calculation of the distance of each object from the reference object, a so-called pattern of development. The pattern (ideal) object, is an abstract object with the best features, i.e., minimum values for cost criteria and maximum values for benefit criteria.

The classical Hellwig method [20] has been modified for real data [40], fuzzy sets [33], intuitionistic fuzzy sets [26, 27, 30, 39], and interval-valued fuzzy sets [41]. The classical Hellwig method and its modifications have been applied for the analysis of complex socio-economic phenomena, such as circular economy [36], quality of human capital in the EU countries [3], socio-economic region development [33], sustainable development [24, 47], quality of life [27, 30], evaluation negotiation offers [39, 43] among others.

A certain modification of the TMD proposed by Hellwig in 1981 was concerned with including in the formula an object referred to as an anti-pattern of development [21]. The measure has been then adjusted to interval data [11, 54], symbolic interval data [53], and fuzzy intuitionistic data [42]. Recent studies using this measure concerned changes in population aging [12], economic efficiency of manufacturing enterprises [11, 53], and social cohesion [54].

In the series of papers, the synthetic measure based on a fuzzy framework was proposed for the analysis of questionnaire surveys with ordinal data [26, 27, 41, 42]. The intuitionistic fuzzy synthetic measure (IFSM) based on distances to the pattern object (ideal solution) was proposed to analyze the survey data with positive, negative, or difficult to say or no opinions [27, 42]. The Euclidean and Hamming distances were applied in the procedure and two pattern object constructions are proposed: one based on maximum and minimum values and the other on maximum intuitionistic fuzzy values. The double intuitionistic fuzzy synthetic measure (DIFSM) inspired by Hellwig's method based on two reference points: the ideal point (pattern) and anti-ideal point (anti-pattern) was proposed in the paper [42]. DIFSM method can handle uncertain, imprecise information or human judgment in survey data evaluation and take into consideration the entropy-based weights of criteria. In another paper, [41] the method of converting ordinal data from questionnaires to interval-valued intuitionistic fuzzy sets and interval-valued intuitionistic fuzzy synthetic measure (I-VIFSM) based on Hellwig's approach were defined. The I-VIFSM measure can be applied for the analysis of complex social phenomena under uncertainty with application data from questionnaire surveys.

In the paper, we modified the Hellwig approach proposed in 1981 to be useful for analyzing complex phenomena where criteria are evaluated by ordinal scale. In the Hellwig procedure, we applied the formula of the GDM2 measure proposed by Walesiak [50] instead of the Euclidean distance measure. This extends the potential application of Hellwig's approach to measuring distances between the object and the pattern of development. We called the proposed measure synthetic measure for ordinal data (SMOD). The SMOD includes steps presented in Figure 3.

Step 1 involved the identification of criteria and objects. Let $O = \{O_1, O_2, \ldots, O_n\}$ be the set of objects and $C = \{C_1, C_2, \ldots, C_m\}$ the set of criteria. There is no restriction on the number of criteria for evaluation.

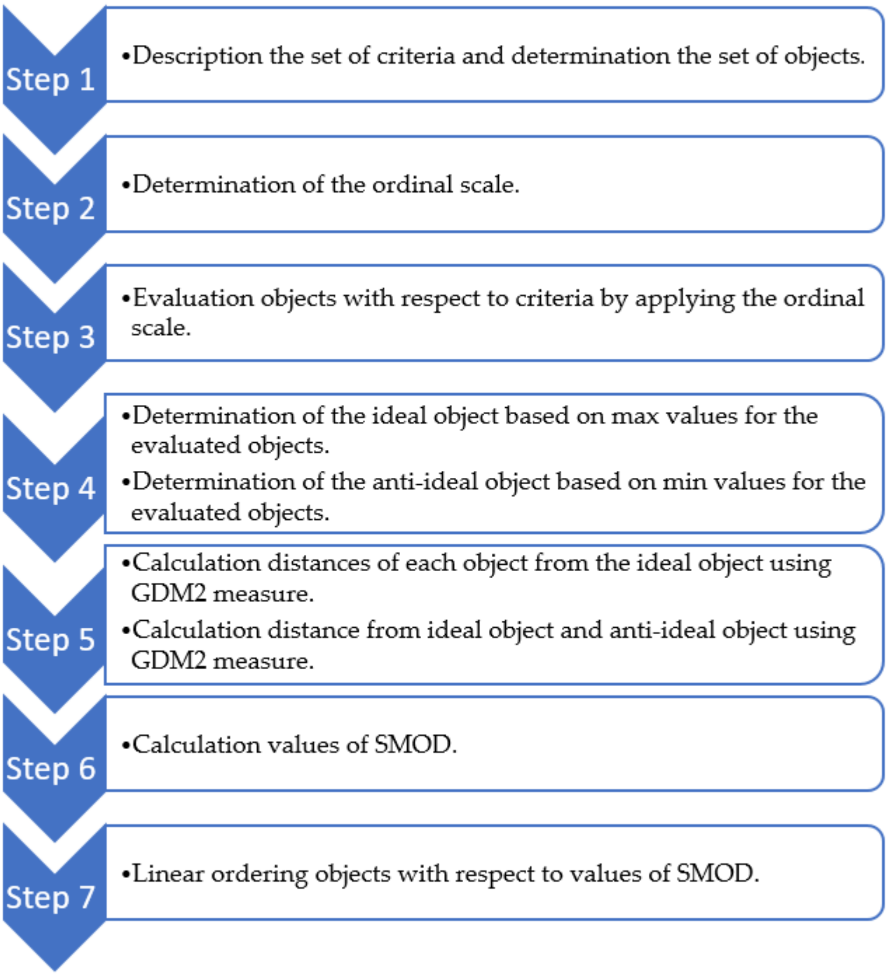

Figure 3. The Steps for Synthetic Measure for Ordinal Data calculation.

In step 2, the ordinal scale is determined. Let $OS = \{1, 2, \ldots, k\}$ be a numerical representation of the ordinal scale. The typical values of cardinality used in the linguistic scale are odd ones, usually between 5 and 13. However, the most popular is the 5-point Likert scale.

In step 3, the evaluation of objects applying the ordinal scale is provided. Let x_{ij} rating of ith object with respect to jth criterion, $x_{ij} \in OS(i = 1, 2, \ldots, n; j = 1, 2, \ldots, m)$. The object O_j $(j = 1, 2, \ldots, m)$ can be represented as a vector $O_j = [x_{1j}, x_{2j}, \ldots, x_{nj}]$, where $x_{ij} \in OS$.

The ideal object $O^+ = [x_1^+, x_2^+, \ldots, x_m^+]$ is determined following the principle:

$$x_j^+ = \max_{i=1,\ldots,n}(x_{ij}) \tag{3}$$

The anti-ideal object $O^- = [x_1^-, x_2^-, \ldots, x_m^-]$ is determined following the principle:

$$x_j^- = \min_{i=1,\ldots,n}(x_{ij}) \tag{4}$$

Determining the coordinates of ideal and anti-ideal objects based on the maximum and minimum values observed in the research sample is one of the most frequently used methods. However, it should be emphasized that there are also other approaches to determine coordinates. The coordinates may be determined by experts or imposed by decision-makers (e.g., in the case of the levels of sustainable development goals that countries should strive for). The coordinates can also be determined based on the maximum and minimum values for the criteria known to the decision-maker (an example would be criteria assessments using ordinal measurement scales with a finite number of scale points).

In step 5, we used GDM2 measure [50] for calculating distances between objects (see formula (1)). We calculate the distance measure GDM2_i^+ between ith object and ideal object O^+ and distance measure GDM2_i^{+-} between ideal object O^+ and anti-ideal object O^- using formula (1).

In step 6, the value of the SMOD measure for the ith object is calculated according to the formula:

$$\text{SMOD}_i = 1 - \frac{\text{GDM2}_i^+}{\text{GDM2}^{+-}} \tag{5}$$

where GDM2_i^+ – generalized distance measure of the ith object from the ideal object, GDM2^{+-} – generalized distance measure between the ideal and the anti-ideal objects, $i = 1, 2, \ldots, n$. The values of SMOD measure are standardized in the range $[0; 1]$.

In step 7, the objects are ordered according to the SMOD values, where the higher the SMOD value, the higher the ranking of the object.

This SMOD can be an alternative or complement to presented earlier methods, i.e., IFSM, DIFSM, and I-VIFSM based on a fuzzy approach. All these methods together give the researchers and practitioner a set of useful tools that can be used for analyzing survey data depending on the problem under consideration and available data. In the SMOD construction, you can use the publicly available `clusterSim` package of the R program. The package includes the `dist.GDM` function, which allows you to significantly speed up step 5 of the SMOD algorithm (see Figure 3).

3. Results

The SMOD method was used in the analysis of the airline service quality performed on TripAdvisor surveys for Star Alliance members. We provided quantitive analysis based on an online survey [see https://www.tripadvisor.com/Airlines accessed on 14.01.2023]. Using data from TripAdvisor or similar user review platforms involves potential biases and limitations. Travelers with extremely positive or negative experiences are more likely to leave reviews, potentially skewing the data. They are typically more motivated to share their opinions, which can lead to self-selection bias. Therefore, these users cannot be considered a random sample reflecting the structure of the general population. Consequently, it is not possible to generalize the findings from a sample to the entire population.

The language in which a review is written can also present challenges, particularly in the context of limitations in data access and usage for global or multilingual analyses. Furthermore, it's worth noting that cultural differences not only influence the way reviews are written but can also highlight variations in the importance attributed to specific product or service attributes by different users. The credibility of the data is further affected by the way opinions are expressed (e.g., concise vs. detailed) and the timing. Reviews left by users at different periods are not directly comparable. They can also be fake and posted, for example, by competing companies or users with an agenda to deliberately denigrate a particular manufacturer or service provider. In this article, all reviews available on the TripAdvisor website were used. Detailed verbal reviews were not taken into account, only the graphic assessment of nine criteria characterizing the airlines and the corresponding rating scale from 1 to 5 (1 – terrible, 2 – poor, 3 – average, 4 – very good, and 5 – excellent). Aggregated data at the level of all users of a given airline were analyzed. TripAdvisor does not share information about the age, gender, education, or other socio-demographic variables of the passengers who wrote the review. Therefore, it is not possible to determine the sample's structure based on these variables.

A total of 376,772 reviews concerning passengers' satisfaction from Star Alliance were completed on TripAdvisor's website. The minimum number of reviews (298) was obtained for Shenzhen Airlines, while the max (52,875) was obtained for United Airlines. Nine questions of the review concerning passenger satisfaction were applied with a five-point measurement Likert scale: 5 – excellent, 4 – very good, 3 – average, 2 - poor, and 1 – terrible. The distribution of indications into individual categories and average scores are presented in Table 1.

Table 1. Overall ratings of airlines.

No.	Airline	Number of interviewers	Linguistic evaluation with a 5-point Likert scale					Average score
			Excelent (5)	Very good (4)	Average (3)	Poor (2)	Terrible (1)	
1	AEGEAN	12,700	6200	3466	987	445	1602	3.962
2	Air Canada	28,554	6025	7454	5071	3261	6743	3.097
3	Air China	5964	872	1677	1429	775	1211	3.038
4	Air India	8554	1554	2276	1608	974	2142	3.015
5	Air New Zealand	13,707	7805	3316	1231	613	742	4.228
6	ANA (All Nippon Airways)	10,365	5333	3251	1009	377	395	4.230
7	Asiana Airlines	3593	1459	1224	515	165	230	3.979
8	Austrian Airlines	8862	2918	2568	1153	658	1565	3.521
9	Avianca	19,678	7067	6283	2312	1152	2864	3.688
10	Brussels Airlines	7921	1674	2707	1331	693	1516	3.294
11	Copa Airlines	13,010	4232	4200	2135	928	1515	3.669
12	Croatia Airlines	1127	260	357	209	111	190	3.343
13	EGYPTAIR	4290	941	1249	874	391	835	3.249
14	Ethiopian Airlines	6040	1302	1903	1125	524	1186	3.267
15	EVA Air	6916	3666	2162	553	223	312	4.250
16	LOT Polish Airlines	5756	1174	1440	931	597	1614	2.994
17	Lufthansa	39,042	13,501	11,522	4872	2744	6403	3.588
18	SAS	10,221	2984	3310	1782	760	1385	3.562
19	Shenzhen Airlines	298	56	107	75	30	30	3.433
20	Singapore Airlines	25,804	15,111	6008	2391	984	1310	4.264
21	South African Airways	4866	1165	1610	991	493	607	3.459
22	Swiss International Air Lines	14754	5564	4235	1845	1061	2049	3.692
23	TAP Air Portugal	22,855	4660	6018	3242	2128	6807	2.982
24	Thai Airways	15,964	6175	5494	2321	896	1078	3.927
25	Turkish Airlines	33,056	13,852	8493	3100	1788	5823	3.689
26	United Airlines	52,875	11,024	14,069	10,687	6749	10,346	3.164

The data from TripAdvisor. Available online: https://www.tripadvisor.com/Airlines (accessed on 14.01.2023).

Overall ratings distribution differ among airlines. Using a linguistic scale to evaluate satisfaction with Star Alliance members, 33.6% of passengers evaluated them as excellent, 28.5% as very good, 14.3% as good, 7.8% as poor, and 16.1% as terrible (Table 1). The number of passengers' opinions for 5 categories is available only for the general assessment of satisfaction from the airline (columns 3–8). Column 9 presents the average score of the general assessment airlines. In the calculation, the linguistic values from excellent to terrible were represented by numerical etiquettes from 5 to 1. The maximum average score of 4.264 was obtained for Singapore Airlines, while the minimum one of 2.982 was obtained for TAP Air Portugal.

Steps 1 and 2. Let us note, that the individual passenger's opinions are provided in terms of rating of consumers' experience with the airline in nine aspects defined earlier. Next, the aggregated opinion is presented in the form of a colored circle which can be transformed into numerical values from a scale from 1 to 5 with inter-middle values $0.5, 1.5, 2.5, 3.5, 4.5$.

Step 3. The airline's rating with respect to all criteria and general rating presented on the TripAdvisor website are collected in Table 2.

Table 2. The airline's rating with respect to all criteria, sum of ratings, and general rating from the TripAdvisor website.

No.	Airline	C1	C2	C3	C4	C5	C6	C7	C8	C9	Sum of ratings	General rating
1	AEGEAN	4	4	3.5	4	4	4	4.5	4	4	36	4
2	Air Canada	3.5	3.5	3.5	3	3.5	3	3.5	3.5	3	30	3
3	Air China	3	3	2.5	3	3	3.5	3.5	3.5	3	28	3
4	Air India	3.5	3.5	2.5	3	3	3.5	3.5	3.5	3	29	3
5	Air New Zealand	4	4	4	4	4.5	4	4.5	4.5	4	37.5	4
6	ANA (All Nippon Airways)	4	4	4	4	4.5	4	4.5	4	4	37	4
7	Asiana Airlines	4	4	3.5	4	4	4	4	4	4	35.5	4
8	Austrian Airlines	3.5	3.5	3	3.5	4	3.5	4	4	3.5	32.5	3.5
9	Avianca	3.5	3.5	4	3.5	4	3.5	4	4	3.5	33.5	3.5
10	Brussels Airlines	3.5	3.5	2.5	3	3.5	3.5	4	3.5	3	30	3.5
11	Copa Airlines	3.5	3.5	3.5	3.5	4	3.5	4	4	3.5	33	3.5
12	Croatia Airlines	3.5	3.5	2	3.5	3.5	3.5	4	3.5	3.5	30.5	3.5
13	EGYPTAIR	3.5	3.5	3	3.5	3.5	3.5	3.5	3.5	3	30.5	3
14	Ethiopian Airlines	3.5	3.5	3	3	3.5	3.5	3.5	3.5	3.5	30.5	3.5
15	EVA Air	4	4	4	4	4.5	4	4.5	4.5	4	37.5	4.5
16	LOT Polish Airlines	3.5	3.5	2.5	3	3.5	3.5	4	3.5	3	30	3
17	Lufthansa	3.5	3.5	3.5	3.5	4	3.5	4	4	3.5	33	3.5
18	SAS	3.5	3.5	2.5	3.5	3.5	3.5	4	4	3	31	3.5
19	Shenzhen Airlines	3.5	3.5	2.5	3	3.5	3.5	3.5	3.5	3	29.5	3.5
20	Singapore Airlines	4	4	4.5	4	4.5	4	4.5	4.5	4	38	4.5
21	South African Airways	3.5	3.5	3	3	3.5	3.5	4	3.5	3.5	31	3.5
22	Swiss International Air Lines	3.5	3.5	3.5	3.5	4	3.5	4.5	4	4	34	3.5
23	TAP Air Portugal	3	3	2.5	3	3.5	3	4	3.5	3	28.5	3
24	Thai Airways	4	4	3.5	4	4	4	4	4	4	35.5	4
25	Turkish Airlines	3.5	4	4	4	4	4	4	4	4	35.5	3.5
26	United Airlines	3.5	3	3	3	3.5	3	3.5	3.5	3	29	3

The data from TripAdvisor. Available online: https://www.tripadvisor.com/Airlines (accessed on 14.01.2023).

Comparing the average score (Table 1, column 9) with the general rating presented graphically in the form of a colored circle on the website (Table 2, column 13) relatively large discrepancies in ratings can be noticed. The two airlines AEGEAN and ANA, for example, are equally positioned in terms of overall rating because they received a general rating of 4. At the same time, the average score for them is 3.962 and 4.23, respectively. Such a situation can happen also for each of the criteria. This can be a source of uncertainty in evaluating airlines while using a multi-criteria approach. The distribution of numerical equivalents and graphical representation scores for the criteria C1-C9 are presented in Table 3.

Airline passengers rated each criterion from poor (2 score) to more than very good (4.5 score). Analyzing the detailed distribution of scores for nine criteria, it can be seen that only in the case of criterion

C3 – in-flight entertainment (Wi-Fi, TV, movies) there are low scores, i.e., 2 and 2.5 scores. When assessing this criterion, airline passengers were also the most divided in their subjective assessments of service quality. In the case of other criteria, the evaluating airlines did not award ratings lower than 3 (average). The highest rated criteria were C7 – cleanliness and C8 – check-in and boarding. Passengers did not give extreme ratings to any criterion, and ratings above the average, i.e., 3.5 over 4 scores, were the most frequently chosen.

Table 3. Airline scores distribution for partial criteria

Score	C1	C2	C3	C4	C5	C6	C7	C8	C9
2	0	0	1	0	0	0	0	0	0
2.5	0	0	7	0	0	0	0	0	0
3	2	3	5	10	2	3	0	0	10
3.5	17	15	7	8	11	15	7	12	7
4	7	8	5	8	9	8	13	11	9
4.5	0	0	1	0	4	0	6	3	0
Total	26	26	26	26	26	26	26	26	26

The data from TripAdvisor.Available online:
https://www.tripadvisor.com/Airlines
(accessed on 14.01.2023).

Step 4. The coordinates of the ideal object are given in Table 4, while the anti-ideal object is in Table 5.

Table 4. The coordinates of the ideal object.

Ideal	C1	C2	C3	C4	C5	C6	C7	C8	C9
Score rating	4	4	4.5	4	4.5	4	4.5	4.5	4

Let us observe that Singapore Airlines are the best airline across all criteria with a rating of 4 for criteria C1, C2, C4, C6, and C9 and 4.5 for other criteria.

Table 5. The coordinates of the anti-ideal object.

Anti-ideal	C1	C2	C3	C4	C5	C6	C7	C8	C9
Score rating	3	3	2	3	3	3	3.5	3.5	3

The closest anti-ideal object is Air China. Higher grades (half scoring points) were obtained only for criteria C3 and C6.

Steps from 5 to 7. We assigned equal weights for all criteria. Using the distance measure for ordinal data (equation (1)) the distances between each airline and the ideal object were calculated and the SMOD values were determined. The results of the calculation distance from airlines to the ideal object, SMOD values, and rank are presented in Table 6. The first with value 1, as an ideal object is ranked Singapore Airlines. The second one is EVA airline which differs from Singapore Airlines in half scoring point only for criterion C3.

The first with SMOD value 1, as an ideal object, is ranked Singapore Airlines. The second one is EVA Air and Air New Zealand. Both airlines differ from Singapore Airlines in half scoring points only for criterion C3. Therefore, both airlines are the same distance from the ideal object, the values of the synthetic measure are the same, and in the ranking, both airlines take second place ex aequo. A similar situation, i.e., obtaining the same result in terms of the distance from the ideal object and the value of the synthetic measure, can also be observed for the following pairs of airlines:

- Asiana Airlines and Thai Airways with rank 6,
- Copa Airlines and Lufthansa with rank 11,
- Brussels Airlines and LOT Polish Airlines with rank 19.

The last is Air China with SMOD value of 0.056. The Pearson coefficient between the average score and SMOD value is equal to 0.956.

Table 6. The distance from the airline and the ideal object, SMOD value, and rank.

Number	Airline	Distance from the ideal object	SMOD	Rank
1	AEGEAN	0.093	0.893	5
2	Air Canada	0.673	0.221	21
3	Air China	0.818	0.053	26
4	Air India	0.733	0.152	23
5	Air New Zealand	0.013	0.985	2
6	ANA (All Nippon Airways)	0.037	0.957	4
7	Asiana Airlines	0.126	0.854	6
8	Austrian Airlines	0.354	0.591	13
9	Avianca	0.279	0.677	10
10	Brussels Airlines	0.639	0.260	19
11	Copa Airlines	0.315	0.636	11
12	Croatia Airlines	0.550	0.364	16
13	EGYPTAIR	0.603	0.302	17
14	Ethiopian Airlines	0.606	0.299	18
15	EVA Air	0.013	0.985	2
16	LOT Polish Airlines	0.639	0.260	19
17	Lufthansa	0.315	0.636	11
18	SAS	0.506	0.414	14
19	Shenzhen Airlines	0.699	0.191	22
20	Singapore Airlines	0.000	1.000	1
21	South African Airways	0.544	0.370	15
22	Swiss International Air Lines	0.248	0.712	9
23	TAP Air Portugal	0.778	0.099	25
24	Thai Airways	0.126	0.854	6
25	Turkish Airlines	0.136	0.843	8
26	United Airlines	0.752	0.129	24

We classified airlines concerning general ratings into four classes. Next, we analyzed average scores (Table 1) and the SMOD values (Table 6) in classes. Descriptive statistics for the average scores and SMOD values with respect to classes are presented in Table 7.

An analysis of the overall level of air passenger satisfaction on TripAdvisor allowed the distinction of four classes of carriers. Class 1 consists of seven airlines, which were rated at the level of average (general rating of 3 scores). The range between the highest and lowest average score rating given by passengers in this group is 0.267. The mean for the average score hovers around 3.077, while the standard deviation is around 0.092. The group can therefore be considered relatively homogeneous in terms of the airline's overall assessment of the quality of passenger service. The same airlines tend to be more diverse when we analyze aggregated passenger opinions on nine service quality criteria. The mean and standard deviation for the synthetic measure SMOD confirm the average differentiation of airlines within the first group, the highest compared to the other three groups.

Table 7. Descriptive statistics for the Average scores and SMOD values with respect to a general rating in classes.

Parameter	Average score	SMOD	Airlines
Class 1. Score 3 (n = 7)			
Mean	3.077	0.174	Air Canada, Air China, Air India, EGYPTAIR, LOT Polish Airlines, TAP Air Portugal, United Airlines
Standard deviation	0.092	0.084	
Min	2.982	0.053	
Max	3.249	0.302	
Class 2. Score 3.5 (n = 12)			
Mean	3.517	0.499	Austrian Airlines, Avianca, Brussels Airlines, Copa Airlines, Croatia Airlines, Ethiopian Airlines, Lufthansa, SAS, Shenzhen Airlines, South African Airways, Swiss International Air Lines, Turkish Airlines
Standard deviation	0.151	0.199	
Min	3.267	0.191	
Max	3.688	0.843	
Class 3. Score 4 (n = 5)			
Mean	4.065	0.909	AEGEAN, Air New Zealand, ANA (All Nippon Airways), Asiana Airlines, Thai Airways
Standard deviation	0.135	0.054	
Min	3.927	0.854	
Max	4.230	0.985	
Class 4. Score 4.5 (n = 2)			
Mean	4.257	0.993	EVA Air, Singapore Airlines
Standard deviation	0.007	0.007	
Min	4.250	0.985	
Max	4.264	1.000	

The next, the largest group (n = 12) are airlines belonging to the second class. According to TripAdvisor users, these are airlines positioned at 3.5 scores, with a maximum average score rating of 3.688 and a minimum rating of 3.267. The mean rating for this group oscillates around 3.517. The airlines forming the second group can therefore be considered relatively homogeneous in terms of the overall level of passenger satisfaction with the quality of service offered by the airlines concerned. However, if we look at the detailed assessments of each of the nine criteria and calculate a synthetic measure SMOD for them, it turns out that the second group is moderately differentiated. Both the first and the second groups have much greater variation in terms of SMOD than the overall rating of passenger satisfaction on TripAdvisor. Therefore, the synthetic measure used in the article has a very important advantage – it is characterized by a high discriminatory capacity.

The range of SMOD values in the second group varies from 0.191 (min) to 0.843 (max). This group of airlines included Shenzhen Airlines, which in terms of synthetic measure obtained a lower position in the ranking than EGYPTAIR from the first group, i.e., that Shenzhen Airlines was separated by a greater distance from the ideal object (0.699) than the one observed for the EGYPTAIR (0.603). This shows discrepancies between the overall assessment and the specific assessments for the nine identified criteria based on which the SMOD value was calculated. EGYPTAIR received higher detailed ratings from passengers compared to Shenzhen Airlines and at the same time a lower overall rating. In terms of SMOD value, EGYPTAIR should be positioned higher than Shenzhen Airlines.

Class 3 consists of 5 airlines positioned at 4 scores, with a maximum average score rating of 4.230 and a minimum rating of 3.927. The mean rating for this group oscillates around 4.065. The range of SMOD values varies from 0.854 (min) to 0.985 (max).

Finally, the last group (Class 4) consists of two of the best airlines, i.e., EVA Air, and Singapore Airlines rating with 4.5 scores.

Air New Zealand from third class and EVA Air from fourth class have identical specific scores for nine criteria, but different overall scores. An overall score of 4 (very good) determined that Air New Zealand belonged to the third group, while the higher rating for EVA Air (general rating of 4.5 scores) determined that it belonged to the fourth group. It can therefore be assumed that EGYPTAIR and Air New Zealand passengers took into account additional and/or different criteria than those included in TripAdvisor in their overall satisfaction assessment. In other words, the overall assessment does not coincide with the specific assessments that were taken into account in the calculation of the SMOD. For both EGYPTAIR and Air New Zealand, the overall ratings on TripAdvisor appear to be underestimated relative to the specific ratings. The question is, which of the assessments – general or aggregated on the basis of nine criteria – is more reliable? It seems that the overall assessment presented in the form of a single numerical value will be, compared to nine specific criteria, more useful for website users and more willingly chosen, for example, as a filter for searching for airlines.

To show the usability of the SMOD approach we can also compare results obtained by the sum of rating. The Pearson coefficient between the sum of ratings (Table 2, column 12) and global rating (Table 2, column 13) equals 0.88895. while the Pearson coefficient between the sum of ratings and SMOD value is 0.99236. The sum of ratings varied from 28 for Air China to 38 for Singapore Airlines. Moreover, for the global rating 3, the sum of ratings varied from 28 (Air China) to 30.5 (EGYPTAIR), 3.5 from 29.5 (Shenzhen Airlines) to 35.5 (Turkish Airlines), 4 from 35.5 (Asiana Airlines, Thai Airways) to 37.5 (Air New Zealand) and 4.5 from 37.5 (EVA Air) to 38 (Singapore Airlines). Additionally, while comparing airlines concerning the sum of ratings and global rating we observed for Shenzhen Airlines global rating 3 and the sum of ratings 29.9, while for EGYPTAIR the global rating is 3 and the sum of ratings 30.5.

The comparison of the results between the SMOD values, sum of ratings for all criteria, and general rating from the TripAdvisor website shows the usefulness of the proposed multi-criteria SMOD approach for analyzing survey web data. The assessment aggregated to the form of SMOD provides more accurate, reliable information and additionally takes into account the uncertainty of the measurement. In addition, like the overall satisfaction rating, it is presented in the form of a single value but synthesizes information for nine specific criteria. It is therefore an interesting proposal to replace detailed assessments with one value of a synthetic measure, which can be competitive to the overall rating posted on TripAdvisor.

The results of rankings based on the SMOD method can be practically utilized by decision-makers and companies in the airline industry. Analyzing consumer opinions regarding specific aspects of the services provided can help airlines identify areas where they excel compared to the competition and areas that require improvement. For example, if consumer opinions indicate low ratings for customer service, airlines can take actions focused on enhancing passenger service quality (e.g., through employee training or acknowledging their dedication to customer service). Analyzing customer needs and expectations enables not only airlines but also businesses in various industries to better align their products and services with consumer expectations. Customer-oriented businesses can gain a competitive advantage, consequently building a base of satisfied and loyal customers.

Airlines that rank high in these rankings can leverage this information in their advertising campaigns, targeting specific passengers seeking particular attributes or high service quality indicators. Promotional materials and advertising can encourage potential customers to choose a specific, highly-rated airline. Moreover, the results of the rankings can be practically used when determining competitive strategies. If a competitive airline consistently receives higher ratings, for example in terms of cleanliness, an airline can take specific actions to improve its score in that area. This ensures it doesn't fall behind the competition while facilitating efficient resource allocation. In other words, efforts are directed where they are needed most. Ranking results can also be used in pricing strategies. Airlines with high positions in the rankings can justify higher prices for their services based on consumer ratings.

Let us note that however, TripAdvisor is recognized as an important information source among users for travel planning some reviews could be biased and misleading. In general, online reviews on the Web may be subject to various kinds of biases [2]. One of them is the self-selection bias reported in user ratings and reviews. It is caused by people's subjective participation in rating or writing reviews online [7, 32, 56]. It is a tendency to give feedback to people with extreme experience, i.e., when they are extremely satisfied or unsatisfied with the product. This could result in a less representative online rating review.

Furthermore, it is worth mentioning that researchers, airlines, and decision-makers should use the findings of these rankings responsibly and ethically. The results of the rankings can be used for good purposes, such as improving service quality, enhancing customer experiences, and promoting healthy competition, rather than, for example, tarnishing competitors.

4. Discussion and conclusions

The development of e-commerce and social media has made online opinions the basic source of information for customers intending to decide to purchase a product/service. E-commerce websites and social media allow customers to introduce comments about products or services by filling out online consumer surveys. Such online opinions called also electronic word-of-mouth have a strong impact on customers' purchasing decisions and are important for business functioning.

To make it easier for consumers to evaluate products/services online, websites offer the use of various types of graphic scales (stars, smiles). From the point of view of measurement theory, these are ordinal scales that are characterized by a limited number of permissible transformations. Therefore, proposals for various types of rankings available on websites and/or social media based on this type of information require some methodological and interpretative caution.

This study has threefold contributions for researchers and practitioners. First, the method for evaluating objects represented by ordinal data by applying the Synthetic Measure for Ordinal Data based on the Hellwig approach is proposed. Second, the usefulness of the SMOD method is described in the context of analysis of complex problems based on website reviews. Finally, this method was applied for rank ordering airline service quality for Star Alliance members. The advantage of the SMOD method is that it handles incomplete information from the online questionnaire reported in the graphical form on TripAdvisor's website.

The SMOD measure proposed by us allows one to evaluate and create rankings of products/services based on the ordinal scales used in online opinions. The measure uses a dedicated GDM2 distance for ordinal data, which protects researchers against the need to artificially strengthen ordinal measurement scales and treat them as metric scales (interval and/or quotient). Obtaining a general rating for airlines requires assigning numbers to individual categories and then averaging the results. This means assuming that the distances between the categories of the scale are the same, which is rarely met. The application of our approach does not require such a procedure, because the GDM2 distance allows only majority and minority relations between the categories of the scale. In addition, the SMOD design does not assume the aggregation of partial criteria in the form of averaging their values, because it is based on GDM2 distances between the evaluated airlines and the ideal object.

The advantage of a multi-criteria approach using SMOD is the assessment of airlines taking into account all sub-criteria. The choice of preferred airlines on the basis of individual criteria is very difficult and ambiguous. In addition, we have shown that it can involve a serious loss of information. The way TripAdvisor rankings are averaged for both overall and sub-criteria makes it difficult to accurately evaluate airlines. Two airlines may have an overall rating of 4, while the average score for these airlines may differ by a value close to 0.5. If such differences appear in the assessments of several sub-criteria, the final ranking of airlines may be strongly distorted. The method we propose makes it possible to build a ranking of airlines based on such imprecise and incomplete information. The proposed SMOD framework can be used both by professionals to improve the quality of functioning airlines and for potential passengers when they have special requirements while choosing the airlines. The ranking results can be valuable for airlines in the context of improving the quality of their services. Information on how passengers rate various aspects of the services provided can help identify both areas where an airline excels and those that require improvement. Taking customer feedback into account characterizes customer-centric businesses and their focus on customer needs. Furthermore, the ranking results for airlines can determine the position of a particular airline compared to its main competitors. Such benchmarks can serve as a starting point for developing strategies to gain a competitive advantage. From the travelers' perspective, the ranking provides information that consumers can use to make more informed decisions when choosing airlines. Selecting an airline that meets the consumer's expectations and preferences contributes to improving their overall experience and loyalty.

SMOD, like any other method, also has certain limitations, arising from, for example, incomplete or missing data that can affect the accuracy of rankings. It happens that reviews may be biased, emotion-based, or untrue and, therefore, cannot be considered representative of the entire population. In other words, the quality and credibility of the data can affect the quality of the rankings. It is also essential to remember that consumer preferences and expectations can change over time. It is therefore recommended to regularly update the rankings to reflect current trends and traveler preferences.

The results of the rankings will also depend on the attributes considered for analysis and the weights assigned to them.

There are a few limitations to this study. Some of them future research is planned to address. To our knowledge, our study is one of the first to analyze TripAdvisor reviews using a multi-criteria approach for ranking airlines implemented GDM2 distance measure. Yet, using only one website in this paper can limit our results. It could be interesting to use our framework on other online review websites. Even if different surveys were applied it could be interesting to compare ranking results. Despite these limitations, proposed the SMOD method is a useful tool for rank ordering airline service quality for Star Alliance members based on TripAdvisors data.

It is also worth noting that the study has a quantitative character, there future searchers look to develop data analyses based on combined approaches of text mining, sentiment analysis, and natural language to investigate more deeply the strengths and weaknesses of each airline. From the analysis of bias occurrence, it could be vital to categorize the different types of passenger reviews. For instance, reviews that give a higher rating are often short, while negative reviews often focus on bad experiences.

Finally, future research on the usefulness of synthetic measures in the analysis of online reviews will focus on the application and methodological aspects. In the first case, it is planned to extend the use of synthetic measures to other sectors of the economy, in particular to the hotel industry, as well as to public institutions and territorial units. The application of SMOD in other sectors requires adapting the methodology to the specific characteristics and requirements of those sectors. This adaptation should consider, among other things, industry-specific factors:

- variables measured on an ordinal scale (e.g., for the hotel industry – room cleanliness satisfaction; in the healthcare industry – patient safety perception; or in the food industry – meal quality),
- correlations between individual variables,
- weights for individual variables/attributes,
- changes in consumer preferences and the need to update rankings accordingly,
- benchmarking and quality and safety standards for specific services and products,
- customer segments,
- ethical and privacy issues,
- spatial and cultural differences.

As a part of the methodological work, it is planned to include an additional parameter of synthetic measures – the timeliness of online reviews. Including an additional parameter in the SMOD method can have both positive and negative effects on rankings and the overall utility of the method. The positive effect will be on the current relevance. Incorporating the timeliness of reviews will make the rankings more up-to-date and aligned with the actual quality of services provided by airlines. Moreover, basing the ranking on the latest reviews will result in a more accurate representation of the customer experience. This is particularly important in industries characterized by rapid changes in trends and/or consumer preferences (e.g., technology or the fashion industry). On the other hand, considering the timeliness of online reviews limits the ability to make time-based comparisons, especially in the context of long-term performance assessment of a particular entity. Additionally, it's essential to remember that within a short period, various events (e.g., promotional campaigns) can lead to a sudden influx of positive or negative reviews, which doesn't always reflect a change (improvement or deterioration) in service quality over an extended period. In summary, the overall usefulness of including the timeliness of online reviews as an additional parameter in the SMOD method depends on the specific industry and the objectives of the analysis.

Acknowledgement

This research was funded by grants from Bialystok University of Technology (WZ/WI-IIT/2/2022) and Wroclaw University of Economics and Business.

References

[1] Atalay, K. D., Atalay, B., and Isin, F. B. FIPIA with information entropy: A new hybrid method to assess airline service quality. *Journal of Air Transport Management 76* (2019), 67–77.

[2] Baeza-Yates, R. Bias on the Web. *Communications of the ACM 61*, 6 (2018), 54–61.

[3] Balcerzak, A. P. Multiple-criteria evaluation of quality of human capital in the European Union countries. *Economics and Sociology 9*, 2 (2016), 11–26.

[4] Balcerzak, A. P., and Pietrzak, M. B. *Digital economy in Polish regions. Proposal of measurement via TOPSIS with generalized distance measure GDM*. Institute of Economic Research Working Papers, No. 3/2017, Institute of Economic Research (IER), 2017.

[5] Basfirinci, C., and Mitra, A. A cross cultural investigation of airlines service quality through integration of Servqual and the Kano model. *Journal of Air Transport Management 42* (2015), 239–248.

[6] Bellizzi, M. G., Eboli, L., and Mazzulla, G. Air transport service quality factors: a systematic literature review. *Transportation Research Procedia 45* (2020), 218–225.

[7] Bhole, B., and Hanna, B. The effectiveness of online reviews in the presence of self-selection bias. *Simulation Modelling Practice and Theory 77* (2017), 108–123.

[8] Chang, B.-C., Ku, C.-H., and Le Nguyen, D.-D. Predicting aspect-based sentiment using deep learning and information visualization: The impact of Covid-19 on the airline industry. *Information and Management 59*, 2 (2022), 103587.

[9] Chen, I.-S. A combined MCDM model based on DEMATEL and ANP for the selection of airline service quality improvement criteria: A study based on the Taiwanese airline industry. *Journal of Air Transport Management 57* (2016), 7–18.

[10] Chou, C.-C., Liu, L.-J., Huang, S.-F., Yih, J.-M., and Han, T.-C. An evaluation of airline service quality using the fuzzy weighted SERVQUAL method. *Applied Soft Computing 11*, 2 (2011), 2117–2128.

[11] Dehnel, G., Gołata, E., and Walesiak, M. A comparative analysis of economic efficiency of medium-sized manufacturing enterprises in districts of Wielkopolska province using the hybrid approach with metric and interval-valued data. *Statistics in Transition New Series 20*, 2 (2019), 49–67.

[12] Dehnel, G., Gołata, E., and Walesiak, M. Assessment of changes in population ageing in regions of the V4 countries with application of multidimensional scaling. *Argumenta Oeconomica 1*, 44 (2020), 77–100.

[13] Dmytrów, K. Comparison of several linear ordering methods for selection of locations in order-picking by means of the simulation methods. *Acta Universitatis Lodziensis. Folia Oeconomica 5*, 338 (2018), 81–96.

[14] Eboli, L., Bellizzi, M. G., and Mazzulla, G. A literature review of studies analysing air transport service quality from the passengers' point of view. *Promet - Traffic&Transportation 34*, 2 (2022), 253–269.

[15] Filieri, R. What makes an online consumer review trustworthy? *Annals of Tourism Research 58* (2016), 46–64.

[16] Filieri, R., Galati, F., and Raguseo, E. The impact of service attributes and category on eWOM helpfulness: An investigation of extremely negative and positive ratings using latent semantic analytics and regression analysis. *Computers in Human Behavior 114* (2021), 106527.

[17] Filieri, R., and McLeay, F. E-wom and accommodation: An analysis of the factors that influence travelers' adoption of information from online reviews. *Journal of Travel Research 53*, 1 (2014), 44–57.

[18] Gupta, H. Evaluating service quality of airline industry using hybrid best worst method and VIKOR. *Journal of Air Transport Management 68* (2018), 35–47.

[19] Gursoy, D., Chen, M.-H., and Kim, H. The US airlines relative positioning based on attributes of service quality. *Tourism Management 26*, 1 (2005), 57–67.

[20] Hellwig, Z. Application of the taxonomic method to the typological division of countries according to the level of their development and the resources and structure of qualified personnel. *Przegląd Statystyczny 15*, 4 (1968), 307–326 (in Polish).

[21] Hellwig, Z. Multivariate Comparative Analysis and Its Application in Multi-Criteria Research of Economic Objects. In *Economic and Mathematical Methods and Models in Improving the Management of Socialist Economies*, W. Welfe, Ed., PWE Warszawa, 1981, pp. 46–68 (in Polish).

[22] Herrera, F., and Herrera-Viedma, E. Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems 115*, 1 (2000), 67–82.

[23] Herrera, F., Herrera-Viedma, E., and Martínez, L. A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE Transactions on Fuzzy Systems 16*, 2 (2008), 354–370.

[24] Iwacewicz-Orłowska, A., and Sokołowska, D. Ranking of EU countries in terms of the value of environmental governance indicators in 2010 and 2015. *Ekonomia i Środowisko-Economics and Environment 66*, 3 (2018), 50–62.

[25] Jalilvand, M. R., Esfahani, S. S., and Samiei, N. Electronic word-of-mouth: Challenges and opportunities. *Procedia Computer Science 3* (2011), 42–46.

[26] Jefmański, B. Intuitionistic fuzzy synthetic measure for ordinal data. In *Classification and Data Analysis. Theory and Apllications* (Cham, 2020), K. Jajuga, J. Batóg, and M. Walesiak, Eds., *Studies in Classification, Data Analysis, and Knowledge Organization* series, Springer, pp. 53–72.

[27] Jefmański, B., Roszkowska, E., and Kusterka-Jefmańska, M. Intuitionistic fuzzy synthetic measure on the basis of survey responses and aggregated ordinal data. *Entropy 23*, 12 (2021), 1636.

[28] Kos Koklic, M., Kukar-Kinney, M., and Vegelj, S. An investigation of customer satisfaction with low-cost and full-service airline companies. *Journal of Business Research 80* (2017), 188–196.

[29] Kuo, M.-S. A novel interval-valued fuzzy MCDM method for improving airlines' service quality in Chinese cross-strait airlines. *Transportation Research Part E: Logistics and Transportation Review 47*, 6 (2011), 1177–1193.

[30] Kusterka-Jefmańska, M., Jefmański, B., and Roszkowska, E. Application of the intuitionistic fuzzy synthetic measure in the subjective quality of life measurement based on survey data. In *Modern Classification and Data Analysis. Methodology and Applications to Micro- and Macroeconomic Problems* (Cham, 2022), K. Jajuga, J. Batóg, and M. Walesiak, Eds., *Studies in Classification, Data Analysis, and Knowledge Organization* series, Springer, pp. 243–261.

[31] Li, W., Yu, S., Pei, H., Zhao, C., and Tian, B. A hybrid approach based on fuzzy AHP and 2-tuple fuzzy linguistic method for evaluation in-flight service quality. *Journal of Air Transport Management 60* (2017), 49–64.

[32] Li, X., and Hitt, L. M. Self-selection and information role of online product reviews. *Information Systems Research 19*, 4 (2008), 456–474.

[33] Łuczak, A., and Wysocki, F. Fuzzy multi-criteria Hellwig's method of linear ordering of objects. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu. Taksonomia 14*, 1169 (2007), 330–340 (in Polish).

[34] Łuczak, A., and Wysocki, F. Application of the generalised GDM distance measure and the TOPSIS method to assess the level of socio-economic development of districts in Wielkopolskie voivodship. *Przegląd Statystyczny 59* (2012), 298–311 (in Polish).

[35] Martínez, L., Ruan, D., and Herrera, F. Computing with words in decision support systems: an overview on models and applications, international. *Journal of Computational Intelligence Systems 3* (2010), 382–395.

[36] Mazur-Wierzbicka, E. Towards circular economy—a comparative analysis of the countries of the European union. *Resources 10*, 5 (2021), 49.

[37] Pakdil, F., and Aydin, Ö. Expectations and perceptions in airline services: An analysis using weighted SERVQUAL scores, journal of air transport management. *Journal of Air Transport Management 13*, 4 (2007), 229–237.

[38] Rosário, A., and Raimundo, R. Consumer marketing strategy and e-commerce in the last decade: A literature review. *Journal of Theoretical and Applied Electronic Commerce Research 16*, 7 (2021), 3003–3024.

[39] Roszkowska, E. The intuitionistic fuzzy framework for evaluation and rank ordering the negotiation offers. In *Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation. Proceedings of the INFUS 2021 Conference, held August 24-26, 2021*. Volume 2 (Cham, 2021), C. Kahraman, S. Cebi, S. Onar, B. Oztaysi, A. Tolga, and I. Sari, Eds., vol. 308 of *Lecture Notes in Networks and Systems*, Springer, pp. 58–65.

[40] Roszkowska, E., and Filipowicz-Chomko, M. Measuring sustainable development using an extended Hellwig method: A case study of education. *Social Indicators Research 153*, 1 (2021), 299–322.

[41] Roszkowska, E., and Jefmański, B. Interval-valued intuitionistic fuzzy synthetic measure (I-VIFSM) based on Hellwig's approach in the analysis of survey data. *Mathematics 9*, 3 (2021), 201.

[42] Roszkowska, E., Jefmański, B., and Kusterka-Jefmańska, M. On some extension of intuitionistic fuzzy synthetic measures for two reference points and entropy weights. *Entropy 24*, 8 (2022), 1081.

[43] Roszkowska, E., Wachowicz, T., Filipowicz-Chomko, M., and Łyczkowska-Hanćkowiak, A. The extended linguistic Hellwig's methods based on oriented fuzzy numbers and their application to the evaluation of negotiation offers. *Entropy 24*, 11 (2022), 1617.

[44] Sezgen, E., Mason, K. J., and Mayer, R. Voice of airline passenger: A text mining approach to understand customer satisfaction. *Journal of Air Transport Management 77* (2019), 65–74.

[45] Stamolampros, P., Dousios, D., and Korfiatis, N. Evaluating domestic bias on airline passengers' ratings: The moderating effect of cultural value orientation. *Journal of Hospitality Management 87* (2020), 102466.

[46] Suki, N. M. Passenger satisfaction with airline service quality in Malaysia: A structural equation modeling approach. *Research in Transportation Business and Management 10* (2014), 26–32.

[47] Szymańska, A. Reducing socioeconomic inequalities in the European Union in the context of the 2030 agenda for sustainable development. *Sustainability 13*, 13 (2021), 7409.

[48] Tsafarakis, S., Kokotas, T., and Pantouvakis, A. A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement. *Journal of Air Transport Management 68* (2018), 61–75.

[49] Tsaur, S.-H., Chang, T.-Y., and Yen, C.-H. The evaluation of airline service quality by fuzzy MCDM. *Tourism Management 23*, 2 (2002), 107–115.

[50] Walesiak, M. Distance measure for ordinal data. *Argumenta Oeconomica 2*, 8 (1999), 167–173.

[51] Walesiak, M. The generalised distance measure GDM as a synthetic measure in linear ordering methods. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu. Taksonomia 10*, 988 (2003), pp. 134–144 (in Polish).

[52] Walesiak, M. Linear ordering with generalised distance measure GDM2 for ordinal data and program R *Econometrics. Ekonometria. Advances in Applied Data Analytics 30* (2011), 9–18.

[53] Walesiak, M., and Dehnel, G. Evaluation of economic efficiency of small manufacturing enterprises in districts of Wielkopolska province using interval-valued symbolic data and the hybrid approach. In *Proceedings of the 12th Professor Aleksander Zeliaś*

International Conference on Modelling and Forecasting of Socio-Economic Phenomena (2018), M. Papież and S. Śmiech, Eds., Foundation of the Cracow University of Economics, pp. 563–572.

[54] Walesiak, M., and Dehnel, G. The measurement of social cohesion at province level in Poland using metric and interval-valued data. *Sustainability 12*, 18 (2020), 7664.

[55] Yan, Z., Xing, M., Zhang, D., and Ma, B. EXPRS: An extended pagerank method for product feature extraction from online consumer review. *Information and Management 52*, 7 (2015), 850–858.

[56] Zhu, Q., Lo, L.Y.-H., Xia, M., Chen, Z., and Ma, X. Bias-aware design for informed decisions: Raising awareness of self-selection bias in user ratings and reviews. *Proceedings of the ACM on Human-Computer Interaction 6*, CSCW2 (2022), 496.