European Journal of Operational Research 220 (2012) 762–776

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Estimation, modeling, and aggregation of missing survey data for prioritizing customer voices

Anil Kumar Maddulapalli [a,*], Jian-Bo Yang [b,1], Dong-Ling Xu [b,2]

[a] General Motors R&D, GMTCI, 2nd Floor Creator Bldg, ITPB, Whitefield Road, Bangalore 560066, India
[b] Manchester Business School, The University of Manchester, Manchester M15 6PB, UK

ARTICLE INFO

Article history:
Received 4 January 2011
Accepted 23 January 2012
Available online 18 February 2012

Keywords:
Decision analysis
Evidential reasoning
Nonlinear optimization
Voices of customer
Partial information

ABSTRACT

It is widely acknowledged that understanding and prioritizing the voice of customer is a critical step in new product development. In this work, we propose a novel approach to handle missing and incomplete data while combining information from different surveys for prioritizing customer voices. Our new approach comprises of the following stages: estimating and representing missing and incomplete data; estimating intervals for the criteria used in analyzing data; mapping data on criteria to a common scale; modeling interval data using interval belief structure; and aggregating evidence and ranking customer voices using the interval evidential reasoning algorithm. We demonstrate our approach using a case study from automotive domain with a given criteria hierarchy for analyzing data from three different surveys. We propose new optimization formulations for estimating intervals of the criteria used in our case study and logical yet pragmatic transformation functions for mapping criteria values to a common scale.

1. Introduction

Quality Function Deployment (QFD) is a widely accepted practice for deploying customer needs (referred to as the voices of customer) through product planning, design and manufacturing (Besterfield et al., 2003; Chan and Wu, 2002). An important step in QFD is to prioritize voices of customer in order to allocate resources appropriately. In many industries, customer satisfaction surveys are routinely conducted for identifying what the customers want, what the strengths of products in market are and where improvements should be made for new products. Data from different surveys are often used to identify those voices of customer that should be given higher priority (sometimes referred to as Key Voices of Customer) within the context and constraints of the overall product or service program (Xie et al., 2010; Yang et al., 2011). For this work, we assume that a list of customer voices to analyze has already been created and we focus on prioritizing the voices with missing and incomplete survey data.

Prioritising voices of customer using data from different surveys involves data analysis with suitable criteria for a survey and then aggregation across criteria and surveys. We are interested in surveys that collect ratings on various voices using a fixed point scale (e.g., 5-point scale with 1 being least satisfied and 5 being most satisfied). For evaluating a voice of customer, raw data from the surveys in the form of ratings is converted into criteria using different metrics (e.g., mean rating) depending on the nature of the survey. Typically multiple criteria are used to evaluate a voice and these criteria can be arranged in a hierarchical structure if the surveys are assumed independent. The criteria hierarchy we use for our case study is shown in Fig. 1 (see Section 3 for more details). Looking at the criteria hierarchy in Fig. 1, it is clear that multiple attribute decision making (MADM) methods are well-suited for aggregating survey data and then prioritizing voices of customer.

Many methods have been proposed in the research community for addressing MADM problems, reader can refer to (Olson, 1996; Okudan and Tauhid, 2008) for good reviews. The Evidential Reasoning (ER) approach (Yang and Singh, 1994; Yang and Sen, 1994; Yang, 2001; Yang and Xu, 2002a,b; Yang et al., 2006) is a unique reasoning-based MADM method that has been applied to many areas such as design and product assessment (Yang and Xu, 1998; Chin et al., 2009). In a previous work (Yang et al., 2011), we have argued that the ER approach is well suited for prioritising voices of customer using survey data. In this work, we use recent developments of the ER approach for prioritising voices of customer with missing and incomplete survey data.

Typically surveys are conducted by external or syndicated agencies (e.g., J.D. Power Associates, Consumer Reports) and the manufacturer or service provider has limited scope in influencing the questionnaires. Due to this, many a time a voice is covered only in a subset of surveys used for analysis. If a voice is not covered in a survey, we refer to it as *missing data* for that voice. Even if a

* Corresponding author. Tel.: +91 80 4041 4766.
E-mail addresses: anil.maddulapalli@gm.com, maklegilimen@yahoo.com (A.K. Maddulapalli), jian-bo.yang@mbs.ac.uk (J.-B. Yang), ling.xu@mbs.ac.uk (D.-L. Xu).
[1] Tel.: +44 161 306 3427.
[2] Tel.: +44 161 275 0941.

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2012.01.045

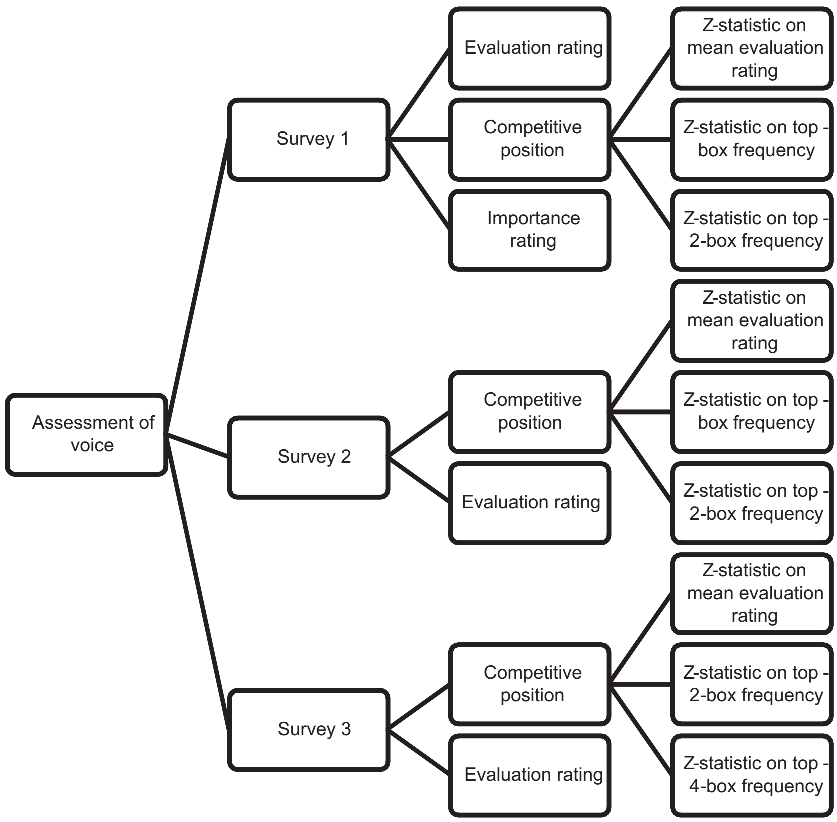

Fig. 1. Criteria hierarchy for evaluating voice of customer in our case study.

voice is covered by a survey, not all respondents give their ratings and this we refer to as *incomplete data* for that voice. In Fig. 1, if there is missing data for a voice, the corresponding survey branch has no information thus leading to imbalance in the way different voices are evaluated. This mismatch in survey coverage for various voices is the motivation for our work. Note that, we use data and information interchangeably throughout this paper.

As the voices of customer for a product are typically related to each other, we wish to exploit those relationships for estimating the missing data for a voice. For each voice, we assume that the set of relevant voices that give good indication about its ratings is known (e.g., a voice hierarchy as shown in Fig. 2). For example, a voice "V1A: Overall, rear seating area roominess meets my needs", can have up to three child voices in a hierarchy, namely "V1A01: Rear seating area has adequate head room", "V1A02: Rear seating area has adequate shoulder room", and "V1A03: Rear seating area has adequate leg room". If the parent voice V1A is not covered by a survey but some or all of the child voices (i.e., V1A01-V1A03) are covered by the same survey or vice-versa, reasonable estimates can be obtained for the voice not covered, using data from voices that are covered (see Section 2.1 for proposed rules of estimation).

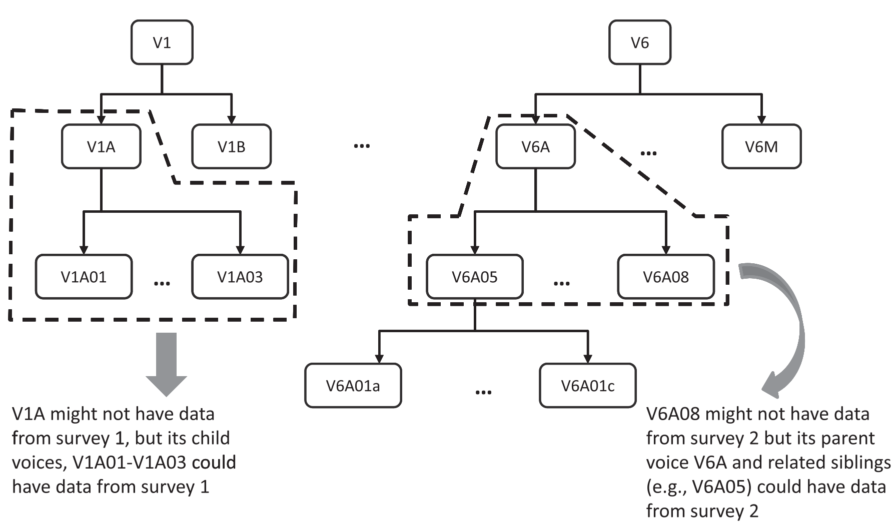

Fig. 2. Example of a voice hierarchy stating relationships between different voices of customer.

Handling missing data can be modelled as decision making under uncertainty or partial information and is well researched in a variety of fields (James and Craig, 1991; Bordley, 1997; Bordley and Kirkwood, 2004; Wallace, 2000; Youk et al., 2004; Manski 2005; Ying et al., 2006; Bordley and Pollock, 2009; Flyer and Hirman, 2009; Liesio and Salo, 2012). Several methods exist in literature for imputing missing responses in a survey (Han and Kamber, 2006). However, when a voice is not covered in a survey, it means that none of the respondents have provided a response, thus raising doubts on the accuracy of imputation methods. In addition, the criteria used in our case study (see Fig. 1) require mean and standard deviation of ratings for a voice, implying that imputation needs to be done at respondent level. When the number of respondents is high, imputation at a respondent level can be computationally very expensive (Bordley, 1997). Also, MADM literature has strongly advocated the use of intervals for quantifying uncertainty in selection problems (Hazen, 1986; Lee et al., 2001; Maddulapalli and Azarm, 2006; Wang et al., 2006a,b; Xu et al., 2006; Sundgren et al., 2009; Park and Jeong, 2011). As our approach is based on MADM, we would like to follow this lead and estimate the missing data using intervals.

In this work, we propose an approach for handling both missing and incomplete survey data in voice of customer analysis, as shown in Fig. 3. In this approach, for a voice, we use available data from relevant voices to estimate the missing survey data as intervals. Incomplete data is handled as ignorance or unassigned evidence within ER algorithm (Yang and Singh, 1994; Wang et al., 2006a,b). We then convert the interval data into intervals for criteria used in evaluating voices (e.g., criteria hierarchy as shown in Fig. 1). As different surveys use different scales, we then transform the interval criteria values to a common scale. Next, we model the interval data on common scale using interval belief degrees, and finally aggregate interval belief degrees using the interval ER algorithm to obtain a utility range for each voice (Wang et al., 2006a,b). The utility ranges for all voices are then used to obtain a partial or complete ranking, depending on the extent of missing and incomplete data for each voice.

The new approach is demonstrated using a case study from automotive domain, selected from a large real life application. In this case study, eight voices with data from up to three surveys are analyzed. These voices have various types of missing and incomplete data. The criteria hierarchy shown in Fig. 1 is used for analyzing these voices. Several new nonlinear optimization formulations are proposed for converting respondent level survey data into intervals for the criteria used shown in Fig. 1. These optimization formulations can be used as guidance for estimating intervals of other such criteria. As different surveys use different scales we also propose intuitive transformation functions to map criteria values to a common scale. The results of our case study show that missing data can be estimated as interval data with confidence, incomplete data can be modelled using a belief structure without loss of any information, Z-statistics can be estimated as intervals for given interval data sets, and robust rankings of voices can be generated using the developed optimization models and process.

Our proposed approach for handling missing and incomplete survey data improves upon our previous works (Xie et al., 2010; Yang et al., 2011) in the following ways. In Yang et al. (2011), survey data is transformed to common scale as crisp numbers by ignoring branches of those surveys (recall Fig. 1) that do not collect information about the voice under evaluation. While (Yang et al., 2011) is one of the first reported works that uses MADM for aggregating survey data to prioritize voices of customers, subsequent experiments using that approach showed that voices with lesser number of data sources have a relatively high chance of topping the priority list. In this work, for maintaining balance in terms of number of information sources for different voices, we estimate missing data using the approach in Fig. 3. Also, we propose new optimization formulations to estimate intervals for the criteria used in evaluating voices and demonstrate aggregation of interval data using the relatively new interval ER algorithm (Wang et al., 2006a,b). On the other hand in Xie et al. (2010) we introduced the concepts of missing data and the need to handle such uncertainty without providing a formal framework to address it. The framework is developed and demonstrated in this work.

The remainder of this paper is organized as follows. In Section 2, we discuss in detail our new approach for handling missing and incomplete survey data in voice of customer analysis. Next in Section 3, we introduce a case study from the automotive domain; propose new optimization formulations for estimating intervals of criteria and functions for transforming criteria values onto the common scale. The results of this case study are discussed in Section 4 and the paper is concluded with a summary in Section 5.

2. A novel approach for handling missing and incomplete survey data

Survey data is usually collected in various forms, e.g., selecting various reasons for purchase by checking boxes, answering yes/no on a statement about a voice, and so on. In this work, we concentrate on data collected using a discrete fixed point scale. Data could be about the importance of a voice in purchase decision or satisfaction/evaluation on a voice statement for a product or service. For example, in our case study (see Section 3 for more details), one of the surveys collects data using a 5-point ordinal scale that is converted into a cardinal scale internally as: "1: DAS, 2: DS, 3: N, 4: A, 5: AS". Respondents are allowed to check only one grade or box of the scale. Respondents would see the grades or boxes "DAS,. . ., AS" in the survey questionnaire and the responses are converted to a 1-5 numeric scale for analysis. Generally one has to be careful in converting ordinal scale data into cardinal data. The conversion is needed if a quantitative metric like Z-statistic

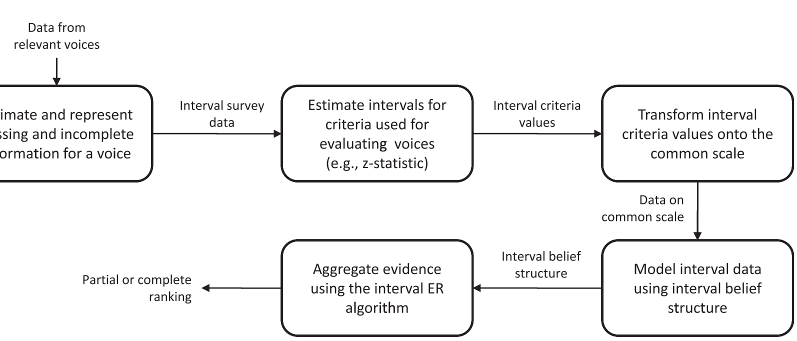

Fig. 3. Schematic of our proposed approach.

or mean rating is used for evaluating voices. Our approach expects the conversion from ordinal scale to cardinal scale as an input and is limited by the assumptions behind that conversion. However if the distribution on the evaluation or importance rating is the only criteria used for assessing voices, then the conversion from ordinal to cardinal scale does not affect our methodology as the ER approach can handle the distribution as it is.

If a voice does not have data from a survey, one could try and estimate an interval for each respondent's response using data on relevant voices (using similar rules that will be discussed in Section 2.1). However, handling interval data at respondent level would be cumbersome and computationally expensive. For example, research has shown that estimating variance of interval data is an NP-hard problem (Ferson et al., 2002). To circumvent issues with handling respondent level interval data, in this research, we work with frequencies of individual grades of the scales or boxes. Using interval data on frequencies is more tractable and this would become clear as we discuss our proposed methods (see Section 3.1).

Many a time, for computational reasons, incomplete responses for a question are completely ignored from analysis, which is not a good practice. Suppose that there are hundred respondents in total for a survey consisting of evaluations on two voices, $V_1$and V_2. Let us say, for V_1 ninety-nine of the respondents checked the same box of the scale (say the first box without loss of generality) and one respondent did not answer, i.e., an incomplete response. For V_2, let us say, ninety-nine of the respondents did not answer, i.e., incomplete responses, and that only one respondent provided an answer by checking the first box of the scale. Now, if the incomplete responses are ignored, one will get the same mean rating (i.e., cardinal value corresponding to the first box of the scale) for V_1 and V_2. However, do we have the same confidence in the mean rating for each of the two voices? The ER approach allows for explicit representation of incomplete information and propagates this information to the top level of criteria hierarchy. At the top level of criteria hierarchy, meaningful rules are used to handle the incomplete information (Yang and Singh, 1994; Yang 2001). If the incomplete responses are taken into account, the frequency of a box (i.e., percentage of total respondents that checked the particular box of the scale) would then be:

$$f_{ijS} = \frac{n_{ijS}}{\sum_{j=0}^{J} n_{ijS}}; \quad \forall j \in \{0, 1, \dots, J\} \tag{1}$$

In Eq. (1), n_{ijs} is the number of respondents[3] that selected the jth box for the ith voice of product S; f_{ijs} is the frequency of the jth box for the ith voice of product S; and J is the number of boxes used in the scale. Also, in Eq. (1), $j = 0$ corresponds to incomplete response and $0 \leqslant f_{ijs} \leqslant 1$. Note that we use the terms "frequency of a box" and "box frequency" interchangeably.

Recall that in this work, we assume that a voice hierarchy stating the relationship between different voices of a product or service is known (see Fig. 2). Using this hierarchy we propose intuitive and logical rules to estimate missing data for voices. In the remainder of this section, we propose methods for the various stages of the approach shown in Fig. 3.

2.1. Stage 1: Estimating and representing missing and incomplete information

After studying through data from different surveys, we represent the information available on a voice from a survey using seven types. We feel that these types are pretty exhaustive, although more surveys could throw light on new types of information. For some of these types of information, box frequencies can be estimated as crisp numbers while for some others, box frequencies have to be estimated as intervals (represented using lower and upper bounds). Next, we discuss the different types of available information and propose rules to estimate and represent missing and incomplete information.

2.1.1. Type 1: Complete information

Data for a given voice is available in the survey of interest and there are no incomplete responses at all. This type has no missing or incomplete information. Calculating frequencies of boxes is straight-forward for this type of information and is given in Eq. (2).

$$f_{ijS} = \frac{n_{ijS}}{\sum_{j=1}^{J} n_{ijS}} \text{ for } j = 1, \dots, J \text{ and } f_{ijS} = 0 \text{ for } \quad j = 0 \tag{2}$$

Note that n_{ijs} is the number of respondents that selected the jth box for the ith voice of product S.

2.1.2. Type 2: Incomplete information

In this type of information, data for a given voice is available in the survey of interest. However there is incomplete data due to lack of responses from some of the respondents. Calculating frequencies of boxes is straight-forward for this type as well and is done using Eq. (1). The difference between Eqs. (1) and (2), is that n_{ijs} and f_{ijs} are non-zero for $j = 0$ (i.e., incomplete response) in Eq. (1).

2.1.3. Type 3: Indirectly complete information

In this type of information, data for a given voice is not directly available in the survey of interest. However, relevant sub-voices (e.g., child nodes of given voice in the voice hierarchy as shown in Fig. 2) of the given voice are all covered by the survey and there are no incomplete responses for the sub-voices. By relevant sub-voices, we mean those child voices in the voice hierarchy, whose survey data can be used to effectively estimate the response for the voice of interest. Let V be the set of relevant sub-voices for the voice of interest and let $|V|$ be the cardinality of the set. The frequency of the jth box for the ith voice of product S, f_{ijs} can then be estimated using

$$f_{ijs} = \sum_{v=1}^{|V|} \alpha_v f_{vjS}; \quad \text{if } j = 1, \dots, J \ \& \ f_{ijs} = 0 \quad \text{if } j = 0;$$
$$0 \leqslant \alpha_v \leqslant 1; \quad \sum_{v=1}^{|V|} \alpha_v = 1 \tag{3}$$

In Eq. (3), f_{vjS} is the frequency of the jth box for the sub-voice $v \in V$ of product S and α_v is the weight associated with sub-voice $v \in V$ in determining the ratings for the given voice. If experience and prior data suggests that a particular sub-voice has more influence in determining the ratings of a given voice, then that sub-voice can be given more weightage by increasing α_v. Otherwise all the sub-voices can be given the same α_v value to denote equal or no different weightage across all sub-voices. If information on $\alpha_v$is not available one could use the minimum and maximum values of f_{vjS} ($v = 1, \dots |V|$) as the lower and upper bound, respectively, of the interval for f_{ijs}.

2.1.4. Type 4: Indirectly incomplete information

In this type of information, a given voice is not covered in the survey of interest, but all relevant sub-voices are covered with incomplete responses. This type is similar to Type 3 but for incomplete responses. For this type of information, we propose to use Eq. (4) for estimating missing and incomplete information.

[3] Some surveys assign weights for respondents based on demographics, sales regions, etc. In such cases, n_{ijs} would be the sum of weights of respondents who selected the jth box for the ith voice of product S.

$$f_{ijs} = \sum_{v=1}^{|V|} \alpha_v f_{vjS}; 0 \leqslant \alpha_v \leqslant 1; \quad \sum_{v=1}^{|V|} \alpha_v = 1; \quad j = 0, \ldots, J \tag{4}$$

2.1.5. Type 5: Indirectly incomplete with partial information

This type of information is more complex to handle than the previous types. In this type of information, a given voice is not covered in the survey of interest and only a few of the relevant sub-voices are covered by the survey with incomplete information. In order to estimate the frequencies of boxes for this type of information, one has to first estimate the frequencies of relevant sub-voices that are not covered by the survey. As before, let V be the set of relevant sub-voices for the voice of interest and let $V' \subset V$ be a subset of V that consists of sub-voices not covered in the survey.

If a sub-voice of a given voice is not covered (directly or indirectly) in a survey, it is logical to conclude that the assessment of the sub-voice could be anything within the survey framework. In other words, the number of responses to any of the boxes and also incomplete responses could be anything between 0% and 100%. An interval of [0,1] is the maximum possible range for f_{ijs}. However, our initial investigations showed that uncertainty in the missing information would be huge, impractical, and intractable if an interval of [0,1] is used. So, we propose to take the minimum and maximum of the frequencies of the covered sub-voices (assuming more than one sub-voice is covered) as the interval for the uncovered sub-voices. Eq. (5) mathematically depicts our proposal for estimating frequency interval of uncovered sub-voice.

$$\begin{aligned} & f_{vjs}^{min} = \min_{u \in V, u \notin V'} \{f_{ujs}\}, \quad f_{vjs}^{max} = \max_{u \in V, u \notin V'} \{f_{ujs}\}; \quad \forall v \in V' \\ & f_{vjs} \in [f_{vjs}^{min}, f_{vjs}^{max}]; \quad \forall j \in \{0, 1, \ldots, J\}; \quad \forall v \in V' \end{aligned} \tag{5}$$

If a sub-voice of a given voice is covered in a survey, the frequencies of boxes for those sub-voices can be obtained using Eq. (6), which is similar to Eq. (1).

$$f_{vjs}(v \in V, v \notin V') = f_{vjs}^{min} = f_{vjs}^{max} = \frac{n_{vjs}}{\sum_{j=0}^{J} n_{vjs}}; \quad \forall j \in \{0, 1, \ldots J\} \tag{6}$$

Using the estimates of the frequencies of sub-voices, the frequency of the jth box for the ith voice of product S can be obtained using Eq. (7), which is similar to Eq. (4).

$$\begin{aligned} & f_{ijS}^{min} = \sum_{v=1}^{|V|} \alpha_v f_{vjS}^{min}; \quad f_{ijS}^{max} = \sum_{(v=1)}^{(|V|)} \alpha_v f_{vjS}^{max}; \quad f_{ijS} \in [f_{ijS}^{min}, f_{ijS}^{max}]; \\ & \forall j \in \{0, 1, \ldots, J\} \end{aligned} \tag{7}$$

In Eq. (7), $j = 0$ would give the frequency of incomplete responses. Note that it is possible that the sub-voices covered by the survey are assessed completely. In such a case, f_{vjS} ($v \notin V'$) would be zero for $j = 0$. It is easy to see that the case of complete information for covered sub-voices is included in the incomplete information case. Note that, Eq. (5) is one possible way of estimating the interval for missing data and other rules can be framed depending on the application, expert knowledge, and so on. We propose to use Eq. (5) based on our experimental results.

2.1.6. Type 6: Indirectly incomplete with partial information for sub-voices

This type of information is also complex to handle. In this type of information, a given voice is not covered by the survey of interest but its parent voice (and none of the sub-voices) is covered by the survey. In order to estimate frequencies for this type of information, one needs to decide the hierarchy for assessing a parent voice and the relative weights (i.e., α_i) of the sub-voices in assessment of the parent voice. Let $\bar{v}$ be the parent voice of a given voice, and $f_{\bar{v}jS}$ be the frequency of the jth box for the parent voice on product S. Let V be the set of relevant sub-voices, including the given voice, that have $\bar{v}$ as the parent, and let $|V|$ be the cardinality of the set. The frequency of the given voice can then be estimated using Eq. (8).

$$\begin{aligned} & f_{ijS} \in [0, f_{ijS}^{max}]; \text{ where } f_{ijS}^{max} = min\{1, f_{\bar{v}jS}/\alpha_i\}; \quad \forall j \\ & \in \{0, 1, \ldots, J\}; \quad 0 \leqslant \alpha_v \leqslant 1; \quad \sum_{v=1}^{|V|} \alpha_v = 1 \end{aligned} \tag{8}$$

If all the sub-voices have same influence on the ratings of the parent voice then $\alpha_i = \frac{1}{|V|}$. As mentioned earlier, experience and prior data should be used in arriving at α_i. Note that the lower bound of f_{ijS} is zero in Eq. (8). The box frequency of a parent voice only provides an upper bound for the children voices' box frequencies as demonstrated by the following example. Assume that there are two sub-voices for a parent voice and that both have an equal weight α_i and that $f_{\bar{v}jS}$ is 0.4. Amongst other possibilities, a value of 0.4 for $f_{\bar{v}jS}$ can be obtained when f_{1jS} has a value of 0.8; f_{2jS} has value of zero and vice versa. Since the only data we have in this type of missing information is on $f_{\bar{v}jS}$ we account for all possible values of f_{ijS} by making its lower bound zero in Eq. (8).

2.1.7. Type 7: Completely missing information

In this type of information, a given voice is not covered in the survey of interest and none of its sub-voices or parent-voice are covered. This type represents total uncertainty about the responses of a voice. In such a scenario, one can only assume that the frequencies can take any permissible values in the whole feasible space as shown in Eq. (9).

$$f_{ijS} \in [f_{ijS}^{min}, f_{ijS}^{max}] = [0, 1]; \quad \forall j \in \{0, 1, \ldots, J\} \tag{9}$$

The seven types of information discussed above cover all kinds of information we have seen in our experimental studies. However, there might be some additional types of information that could be encountered in other data sources. For example, for a given voice not covered in the survey of interest, its parent voice and some of its sub-voices might be covered in the survey with incomplete information. The guidelines established in the above discussion would help in arriving at rules for estimating missing data in most of the situations.

2.2. Stage 2: Estimating intervals for criteria used in analyzing data

Data from surveys are usually analyzed using a variety of statistical and mathematical criteria. Due to missing information, survey data for a voice might be interval in nature. Calculating the criteria used in survey analysis for interval data might not be straight-forward if the underlying metrics are non-linear and complex. If the criteria used for analyzing survey data can be expressed in terms of the frequencies of the boxes used in the scale of the survey questionnaire, it is tractable to estimate intervals of criteria used in evaluating a voice. Optimization formulations can be developed for estimating the minimum and maximum of the criteria value using the intervals of box frequencies as constraints. In Section 3.1, we propose such optimization formulations for the criteria used in our case study.

2.3. Stage 3: Mapping data to a common scale

In the ER approach, a belief decision matrix (Yang, 2001) is used to represent the data on voices. Different surveys use different scales for collecting data and analysts use disparate criteria for analyzing the survey data (Yang et al., 2011). To aggregate data in a uniform way, the ER approach maps the data onto a common

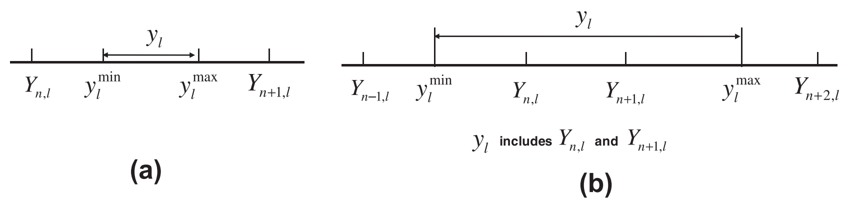

Fig. 4. Schematic of interval data (a) within two adjacent levels (b) including at least one level.

scale and then uses a belief structure on the common scale. The rules and functions used for transforming data to the common scale are problem dependent and so is the common scale.

In general, the common scale contains discrete grades similar to the grades used for eliciting preferences in surveys. For criteria such as evaluation ratings, a mapping is required between the grades used in the survey questionnaire and the common scale grades (e.g., DAS maps to H_1, AS maps to H_5). For criteria that can take continuous values (e.g., mean evaluation rating, Z-statistic), discrete levels need to be identified for mapping to the common scale. For any criterion value between two such identified levels, we can use linear interpolation for calculating the assessment to the common scale grades. For example, let one of the criteria used to evaluate voices be the mean evaluation rating from a survey that collects data using a five-point scale (1 being the worst and 5 being the best). Note that mean rating is continuous and in this example can take any value between one and five. For simplicity assume that mean rating of 1 maps to H_1 grade of common scale, mean rating of 2 maps to H_2, and so on with mean rating of 5 mapping to H_5. For a voice that has a mean rating of 3.4 from this survey, we know that there is evidence towards the common scale grades H_3 and H_4 only. Thus, using linear interpolation we can assign a belief of 0.6 to H_3 and 0.4 to H_4. In Section 3.2, we discuss the specific details of the common scale and the transformation functions that we use for our case study.

2.4. Stage 4: Modelling interval data using interval belief structure

For criteria that take values on discrete grades, it is straight-forward to map the data to common scale even when the values are interval in nature. For criteria that take continuous values, mapping to the common scale needs special attention when the values are interval in nature. Herein we discuss an existing approach (Wang et al., 2006a) for modeling interval data on criteria using an interval belief structure. For the lth criterion of the criteria hierarchy used in evaluating voices, let $Y_{1,l}, \ldots, Y_{n,l}, \ldots, Y_{N,l}$ be the levels of the continuous domain that are used for mapping to the common scale, $\{H_1, \ldots, H_N\}$[4]. Let y_l be a criterion value for which the belief degree $\beta_{n,l}$ on the common scale needs to be calculated. If $y_l$is precise (as is the case in the example discussed at the end of Section 2.3), for $Y_{n,l} \leqslant y_l \leqslant Y_{n+1,l}$, using linear interpolation, we can assign belief degrees to common scale grades H_n and H_{n+1} as $\beta_{n,l} = \frac{Y_{n+1,l}-y_l}{Y_{n+1,l}-Y_{n,l}}$ and $\beta_{n+1,l} = \frac{y_l-Y_{n,l}}{Y_{n+1,l}-Y_{n,l}}$. The belief degree for other grades of common scale would be zero.

If y_l is interval in nature, i.e., $y_1 \in [y_l^{min}, y_l^{max}]$, one can envision two scenarios as shown in Fig. 4. In the first scenario, interval data is completely contained within two adjacent criterion levels used for mapping to common scale (as shown in Fig. 4a) and in the second scenario, interval data includes at least one criterion levels (as shown in Fig. 4b). As the second scenario is more generic, we discuss below the approach proposed in (Wang et al., 2006a) for modelling interval data using an interval belief structure.

Without loss of generality, consider a situation where two criterion levels are included in the interval $y_1 \in [y_l^{min}, y_l^{max}]$. For other situations one can model in the same way. Suppose $Y_{n-1,l}$ and $Y_{n+2,l}$ are the two closest criterion levels that include the interval $y_l \in [y_l^{min}, y_l^{max}]$ as shown in Fig. 4b. Then, y_l can be equivalently expressed in the form of belief structure as (Wang et al., 2006a)

$$y_l \Longleftrightarrow \{(H_{n-1,l}, [\beta_{n-1,l}^{min}, \beta_{n-1,l}^{max}]);\quad (H_{n,l}, [\beta_{n,l}^{min}, \beta_{n,l}^{max}]);\\ (H_{n+1,l}, [\beta_{n+1,l}^{min}, \beta_{n+1,l}^{max}]);\quad (H_{n+2,l}, [\beta_{n+2,l}^{min}, \beta_{n+2,l}^{max}])\} \tag{10a}$$

$$\beta_{n-1,l} + \beta_{n,l} + \beta_{n+1,l} + \beta_{n+2,l} = 1 \text{ and } I_{n-1,n} + I_{n,n+1} + I_{n+1,n+2} = 1 \tag{10b}$$

$$\beta_{n-1,l}^{min} = 0 \text{ and } \beta_{n-1,l}^{max} = I_{n-1,n} \cdot \frac{Y_{n,l} - y_l^{min}}{Y_{n,l} - Y_{n-1,l}} \tag{10c}$$

$$\beta_{n,l}^{min} = 0 \text{ and } \beta_{n,l}^{max} = I_{n-1,n} + I_{n,n+1} \tag{10d}$$

$$\beta_{n+1,l}^{min} = 0 \text{ and } \beta_{n+1,l}^{max} = I_{n,n+1} + I_{n+1,n+2} \tag{10e}$$

$$\beta_{n+2,l}^{min} = 0 \text{ and } \beta_{n+2,l}^{max} = I_{n+1,n+2} \cdot \frac{y_l^{max} - Y_{n+1,l}}{Y_{n+2,l} - Y_{n+1,l}} \tag{10f}$$

and $I_{n-1,n}$; $I_{n,n+1}$; $I_{n+1,n+2}$ are 0–1 binary variables with

$$I_{k-1,k} = \begin{cases} 1, & \text{if } y_l \text{ lies between } \mathrm{Y}_{k-1,l} \text{ and } \mathrm{Y}_{k,l} \\ 0, & \text{otherwise} \qquad k = n, n+1, n+2 \end{cases} \tag{10g}$$

In Eq. (10), binary variables are needed to ensure that for a realisation of y_l from the range $\left[y_l^{min}, y_l^{max}\right]$, only a pair of adjacent criterion levels and the corresponding common scale grades have non-zero belief degrees. Readers can find more details on Eq. (10) in (Wang et al., 2006a). The main reason for reproducing Eq. (10) here is to highlight the fact that because of the binary variables as shown in Eq. (10), the optimization models used in the interval ER algorithm (see Section 2.5) turn out to be Mixed Integer Non-Linear Programming (MINLP) problems.

With the help of the transformation functions and belief structure expressions as shown in Eq. (10), the interval data on the criteria of voice can be transformed to an interval belief structure on the common scale. The next stage in our approach is to aggregate the interval belief degrees on various criteria to come up with an overall utility score for rank ordering various voices

2.5. Evidence aggregation using the interval ER algorithm

The aggregation of the evidence transformed to the common scale needs to be consistently conducted from the bottom level of the criteria hierarchy to higher levels. The aggregation method based on the interval ER approach (Wang et al., 2006a) is briefly described here. Due to interval data, the ith voice, V_i, can be assessed on a criterion c_l using the following interval belief structure $S(c_l(V_i)) = \{(H_n, [\beta_{n,l}^{min}(V_i), \beta_{n,l}^{max}(V_i)]), n = 1, \ldots, N\}$ where H_n is the nth grade in the common scale, $\beta_{n,l}(V_i)$ the belief degree (assess-

[4] For demonstration we are assuming that the number of levels in the lth criterion is the same as the number of grades in the common scale. However these can be different in actual applications.

ment) to which voice V_i is assessed to the grade H_n on the lth criterion, $\beta_{n,l}^{min}(V_i)$ the minimum estimate of $\beta_{n,l}(V_i)$, and $\beta_{n,l}^{max}(V_i)$ the maximum estimate of $\beta_{n,l}^{max}(V_i)$. If the interval belief degrees $\left[\beta_{n,l}^{min}(V_i), \beta_{n,l}^{max}(V_i)\right]$ always satisfy $\sum_{n=1}^{N}\beta_{n,l}(V_i) = 1$ in any circumstances, where $\beta_{n,l}(V_i) \in \left[\beta_{n,l}^{min}(V_i), \beta_{n,l}^{max}(V_i)\right]$ for $n = 1, \ldots, N$, then $S(c_l(V_i))$ is said to be a complete interval distribution assessment; otherwise, it is incomplete. For an incomplete interval distribution assessment, the belief degree, $\beta_{H,l}(V_i) = 1 - \sum_{n=1}^{N}\beta_{n,l}(V_i)$ that could be assigned to any of grades in the common scale is an interval defined by $\left[\beta_{H,l}^{min}(V_i), \beta_{H,l}^{max}(V_i)\right]$.

Suppose there are L criteria on which to assess the voice V_i and w_l be the relative weight of the lth criterion. The overall belief degree for V_i (denoted by $\beta_n(V_i)$ for H_n and $\beta_H(V_i)$ for the unassigned or incomplete belief), is an aggregation of the belief degrees on the individual criteria and can be obtained using the set of equations shown in Eq. (11).

$$\beta_n(V_i) = \frac{m_n(V_i)}{1 - \bar{m}_H(V_i)}; \quad \beta_H(V_i) = \frac{\tilde{m}_H(V_i)}{1 - \bar{m}_H(V_i)}; \quad n = 1, \ldots, N \tag{11a}$$

$$\begin{aligned} m_n(V_i) =& k \cdot \left[\prod_{l=1}^{L}\left(w_l\beta_{n,l}(V_i) + (1 - w_l) + w_l\beta_{H,l}(V_i)\right)\right. \\ &\left. - \prod_{l=1}^{L}\left((1 - w_l) + w_l\beta_{H,l}(V_i)\right)\right]; \quad n = 1, \ldots, N \end{aligned} \tag{11b}$$

$$\widetilde{m}_H = k \cdot \left[\prod_{l=1}^{L}\left((1 - w_l) + w_l \cdot \beta_{H,l}(V_i)\right) - \prod_{l=1}^{L}(1 - w_l)\right] \tag{11c}$$

$$\overline{m}_H = k \cdot \left[\prod_{l=1}^{L}(1 - w_l)\right] \tag{11d}$$

$$\begin{aligned} k = &\left[\sum_{n=1}^{N}\prod_{l=1}^{L}(w_l \cdot \beta_{n,l}(V_i) + (1 - w_l) + w_l \cdot \beta_{H,l}(V_i))\right. \\ &\left. -(N - 1)\prod_{l=1}^{L}((1 - w_l) + w_l \cdot \beta_{H,l}(V_i))\right]^{-1} \end{aligned} \tag{11e}$$

$$\beta_{n,l}^{min}(V_i) \leqslant \beta_{n,l}(V_i) \leqslant \beta_{n,l}^{max}(V_i); \quad n = 1, \ldots, N; \quad l = 1, \ldots, L \tag{11f}$$

$$\beta_{H,l}^{min}(V_i) \leqslant \beta_{H,l}(V_i) \leqslant \beta_{H,l}^{max}(V_i); \quad l = 1, \ldots, L \tag{11g}$$

$$\sum_{n=1}^{N}\beta_{n,l}(V_i) + \beta_{H,l}(V_i) = 1; \quad l = 1, \ldots, L \tag{11h}$$

The aggregation shown in Eq. (11) is commonly referred in the literature as the analytical ER algorithm (Wang et al., 2006b) and is based on the Yen's rule of combination (Yen, 1990), which states that the normalization of combined evidence can be conducted at the end of combination of evidence without changing the combination result. ER framework models ignorance clearly by breaking down unassigned belief degree and the corresponding probability mass into two parts (i.e., $\tilde{m}_H$ and $\bar{m}_H$) and treating them differently, thus avoiding some of the pitfalls of the Dempster–Shafer rule for combining evidence (Wang et al., 2006a). Eqs. (11b), (11c), (11d) are used to combine evidence for belief degrees on the common scale grades and the two parts of the unassigned belief degree. Eq. (11e) is the normalization constant. Eq. (11f) and (11g) are used to ensure that belief degrees are within the specified intervals and Eq. (11h) ensures that belief degrees sum up to one. Note that Eq. (11) represents belief aggregation for one level of the criteria hierarchy only. If the criteria hierarchy has more than one level, then $\beta_{n,l}(V_i)$ and $\beta_{H,l}(V_i)$ would be functions of the belief degrees of the corresponding sub-criteria. More constraints need to be added to Eq. (11) in such cases.

The interval ER algorithm uses non-linear optimization models for estimating the intervals for the overall belief degrees $\beta_n(V_i)$ and $\beta_H(V_i)$. The maximum and minimum values, and hence the interval, for overall belief degrees can be obtained using a non-linear optimization problem with the expression for $\beta_n(V_i)$ or $\beta_H(V_i)$ in Eq. (11) as the objective and other expressions as constraints. Note that, for a 5-point common scale, twelve non-linear optimization problems need to be solved to obtain the lower and upper bounds of the intervals for each grade plus the incomplete response. From here-on, we use the notation $\{\beta_n(V_i), \beta_H(V_i)\} \in \theta(V_i)$ to denote the space that satisfies Eq. (11).

It is possible that the overall belief degree intervals would not result in a consistent rank order because of overlaps in the distributed assessments. For handling such cases, interval ER approach uses minimum, maximum, and average utilities for ranking or prioritization. Let $U(H_n)$ denote the utility for the nth grade of the common scale and without loss of generality assume that utility increases with n. $U(H_n)$ is a cardinal number, usually between zero and one, that reflects the preference of the decision maker and quantifies the value that a voice has when it is assigned to the nth grade of the common scale. The interval ER approach uses Eq. (12) for obtaining the maximum value of the overall utility, $U_{max}(V_i)$, and Eq. (13) for obtaining the minimum, $U_{min}(V_i)$.

$$\text{Maximize} \quad U_{max}(V_i) = \sum_{n=1}^{N}U(H_n) \cdot \beta_n(V_i) + U(H_N) \cdot \beta_H(V_i) \tag{12a}$$

$$\text{Subject to} \quad \{\beta_n(V_i), \beta_H(V_i)\} \in \theta(V_i) \tag{12b}$$

$$\text{Minimize} \quad U_{min}(V_i) = U(H_1) \cdot \beta_H(V_i) + \sum_{n=1}^{N}U(H_n) \cdot \beta_n(V_i) \tag{13a}$$

$$\text{Subject to} \quad \{\beta_n(V_i), \beta_H(V_i)\} \in \theta(V_i) \tag{13b}$$

Note that in Eq. (12) the unassigned overall belief degree, $\beta_H(V_i)$, is assigned to the most preferred grade H_N and in Eq. (13) it is assigned to the least preferred grade H_1. $\beta_H(V_i)$ is positive only when there are incomplete responses in a survey and would be zero otherwise. The beauty of the ER framework is that such ignorance or incompleteness is treated consistently and the question of assigning the incompleteness to a particular grade occurs only after aggregation. Since $\beta_H(V_i)$ can be assigned to any grade of the common scale, it is intuitive that maximum (minimum) utility occurs when it is assigned to the most (least) preferred grade.

Note that the optimization formulations in Eqs. (12) and (13) are non-linear programming (NLP) problems. In many cases, as mentioned in Section 2.4, they turn out to be MINLP problems due to the binary variables that are needed to accurately model the interval belief degree space, $\theta(V_i)$. Even though the optimization formulations in the interval ER approach are non-linear, they can be solved relatively easily using standard optimization packages like LINGO (Wang et al., 2006a). However, in most cases the formulations are non-convex and hence global optimality cannot be guaranteed. Also, the complexity of the optimization formulation increases with the number of criteria, the number of levels in the criteria hierarchy, and the number of grades in the common scale.

Next we discuss a case study for applying our approach and propose optimization formulations and transformation functions for converting missing survey data into belief degree intervals on the common scale.

3. Case study from automotive domain: new nonlinear estimation models and data transformation functions

To demonstrate our approach for handling missing and incomplete survey data, we selected an automotive case study involving ranking of eight voices based on data from three surveys. Two of these surveys are conducted by external agencies and the other one is an internal survey. In Survey 1, respondents are asked to evaluate vehicles on various voices using a 5-point scale with the following grades/boxes: "1: DAS, 2: DS, 3: N, 4: A, 5: AS". The respondents are also asked to state the importance of various voices in their purchase decision on a 5-point scale as follows: "1: LI, 2: NI, 3: I, 4: VI, 5: EI". Survey 2 and Survey 3 ask the respondents to evaluate vehicles on various voices using a 5-point scale ("0: NG, 1: G, 2: VG, 3: E, 4: O") and a 10-point numeric scale respectively. Survey 1 is a focused survey and has a couple of hundreds of carefully pre-screened respondents. Survey 2 and Survey 3 are mail-back surveys and have a few thousands of respondents. The relevant part of the questionnaires in all three surveys consists of around hundred statements. Note that, as we are dealing with frequencies of the grades used in a survey scale, the number of respondents in the survey does not have a bearing on the computational aspects of our proposed approach.

As the three surveys in our case study use different scales, we propose to use the following common scale: "No priority, Low Priority, Average Priority, High Priority, Top Priority" for aggregating data (Yang et al., 2011). The goal is to compare the data on manufacturer's vehicle (in a particular segment) with competition and determine the rank order for prioritizing the voices for resource commitment.

To assess a voice using data from the three surveys on manufacturer's vehicle and competition, various criteria are used as shown in Fig. 1. For all the surveys, the evaluation ratings for the manufacturer's vehicle on a voice are used as a criterion. The distribution on the rating scale from each survey is converted into the common scale using the transformation functions discussed in Section 3.2. Also, for all three surveys competitive position is used as a criterion. Using this criterion manufacturer's vehicle is statistically compared with the competition. Three sub-criteria are used for assessing voices using competitive position. All three sub-criteria use Z-statistic on various metrics of interest. Survey 1 and Survey 2 use Z-statistic on mean evaluation rating, top-box frequency and top-2-box frequency. Top-box frequency refers to the frequency for the most appealing (top) grade of the scale and top-2-box frequency refers to the sum of frequencies of the most and the second-to-most appealing (top 2) grades of the scale. Survey 3 uses Z-statistic on mean evaluation rating, top-2-box frequency, and top-4-box frequency. In addition, Survey 1 uses the distribution on the importance ratings for assessing a voice.

The voices we are dealing with in this case study are the needs that typical customers look for in an automobile. The voices that need to be rank ordered are code named V3B, V3C, V5B08, V6A05, V6A06, V6B05, V6B11 and V6C09. The first two voices deal with the interior and exterior appearance of the vehicle. Fourth and fifth voices deal with ease of getting in and out of various seats of the vehicle. Sixth and seventh voices deal with roominess (e.g., head room, leg room) at various positions in the vehicle. Further descriptions of these voices are not revealed here due to confidentially reasons. Data from the three surveys on the eight voices encompass the seven types of information discussed in Section 2.1. Table 1 shows the type of information that is available for each of these voices from all the three surveys. As some of the voices have missing information and incomplete information, we use our proposed approach (Fig. 3) to estimate the missing information and then aggregate the information for ranking the voices.

Using the approach described in Section 2, we would first estimate the intervals for missing data, then convert the interval data into intervals for criteria, transform the interval criteria values to common scale, model the interval data on common scale using interval belief degrees, and finally aggregate interval belief degrees using the interval ER algorithm. To accomplish these steps, we need methods to convert interval survey data as intervals for criteria used in Fig. 1 and transformation functions to map criteria values to the common scale. These are discussed next.

3.1. Nonlinear optimization models for estimating intervals using frequency data

The criteria used to assess voices in our case study, as shown in Fig. 1, can be broadly classified into three categories: (a) Evaluation/Importance rating, (b) Z-statistic on mean evaluation rating, and (c) Z-statistic on cumulative box frequency. Cumulative box frequency implies sum of the frequencies of the boxes of interest, e.g., top-2-box, top-4-box. Of these categories, interval data on evaluation/importance rating can be directly estimated using the rules described in Section 2.1. For the other two categories, interval data on box frequencies needs to be converted into corresponding criteria intervals. Next, we discuss our proposed method for estimating intervals of Z-statistic on mean evaluation rating when the input survey data is interval in nature.

3.1.1. Estimating intervals for Z-statistic on mean evaluation rating

Eq. (14) gives the mathematical expression for calculating the Z-statistic on the mean evaluation rating of manufacturer's vehicle (S) with respect to the competition (R) for the ith voice, $z_{i\mu}$. In Eq. (14) and elsewhere, N_{iS}, N_{iR} are the total number of respondents; μ_{iS}, μ_{iR} are the mean evaluation ratings; and σ^2_{iS}, σ^2_{iR} are the variances in evaluation ratings.

$$Z_{i\mu} = \frac{\mu_{iS} - \mu_{iR}}{\sqrt{\frac{\sigma^2_{iS}}{N_{iS}} + \frac{\sigma^2_{iR}}{N_{iR}}}} \tag{14}$$

As the survey data we are interested in is collected using fixed scales, the mean evaluation rating and the variance in the evaluation ratings of the ith voice for S can be obtained using Eq. (15).

$$\mu_{iS} = \sum_{j=1}^{J} j \cdot f_{ijS} \quad \text{and } \sigma^2_{iS} = \frac{\sum_{j=1}^{J} n_{ijS}(j-\mu_{is})^2}{(\sum_{j=0}^{J} n_{ijS}) - 1} = \frac{\sum_{j=1}^{J} f_{ijS}(j-\mu_{is})^2}{1 - \frac{1}{\sum_{j=0}^{J} n_{ijS}}} \tag{15}$$

In Eq. (15), n_{ijS} is the number of respondents that selected the jth box for the ith voice of S and f_{ijS} is the the frequency for the jth box for the ith voice of S. Note that $N_{iS} = \sum_{j=0}^{J} n_{ijS}$. The mean and variance of the evaluation ratings for R can be obtained by substituting R by S in Eq. (15). With the above definitions of mean rating, variance of ratings, and Z-statistic on mean rating, we propose the following nonlinear programming model to obtain the interval for Z-statistic on mean rating.

$$\text{Minimize and Maximize } \; Z_{i\mu} = \frac{\mu_{iS} - \mu_{iR}}{\sqrt{\frac{\sigma^2_{iS}}{N_{iS}} + \frac{\sigma^2_{iR}}{N_{iR}}}} \tag{16a}$$

$$\mu_{iS} = \sum_{j=1}^{J} j \cdot f_{ijS} \text{ and } \mu_{iR} = \sum_{j=1}^{J} j \cdot f_{ijR} \tag{16b}$$

$$\sigma^2_{iS} = \frac{\sum_{j=1}^{J} f_{ijS}(j-\mu_{iS})^2}{1 - \frac{1}{\sum_{j=0}^{J} n_{ijS}}} \text{ and } \sigma^2_{iR} = \frac{\sum_{j=1}^{J} f_{ijR}(j-\mu_{iR})^2}{1 - \frac{1}{\sum_{j=0}^{J} n_{ijR}}} \tag{16c}$$

$$\text{Subject to } \; \sum_{j=1}^{J} f_{ijS} + f_{i0S} = 1 \text{ and } \sum_{j=1}^{J} f_{ijR} + f_{i0R} = 1 \tag{16d}$$

$$f_{ijS}^{min} \le f_{ijS} \le f_{ijS}^{max} \text{ and } f_{ijR}^{min} \le f_{ijR} \le f_{ijR}^{max}; \quad \forall j \in \{0, 1, \ldots, J\} \tag{16e}$$

Table 1
Available information types for the eight voices in the case study.

Voice	Survey 1	Survey 2	Survey 3
V3B	Type1	Type 2	Type 2
V3C	Type1	Type 7 {neither parent voice nor sub-voices are surveyed}	Type 2
V5B08	Type 2	Type 2	Type 2
V6A05	Type 3 {sub-voices V6A05a and V6A05b are surveyed with complete information}	Type 2	Type 6 {parent voice V6A is surveyed and sub-voices V6A05a and V6A05b are not surveyed}
V6A06	Type 3 {sub-voices V6A06a and V6A06b are surveyed with complete information)	Type 2	Type 6 {parent voice V6A is surveyed and sub-voices V6A06a and V6A06b are not surveyed}
V6B05	Type 4 {sub-voices V6B05a, V6B05b, and V6B05c are surveyed with incomplete information}	Type 4 {sub-voices V6B05a, V6B05b, and V6B05c are surveyed with incomplete information}	Type 2
V6B11	Type 3 {sub-voices V6B11a, V6B11b, and V6B11c are surveyed with complete information}	Type 5 {sub-voice V6B11a and V6B11c are surveyed with incomplete information but V6B11b is not surveyed}	Type 2
V6C09	Type 5 {sub-voice V6C09a and V6C09c are surveyed with incomplete information but V6C09b is not surveyed}	Type 2	Type 2

In Eq. (16), objective function is the Z-statistic of the mean evaluation rating as defined in Eq. (14). Eqs. (16b) and (16c) are used to constrain the mean values and variances of S and R according to Eq. (15). Eq. (16d) ensures that the sum of frequencies for S and R add up to one. Eq. (16e) restricts the allowed values for the frequencies of S and R to the intervals obtained using Eqs. (1)–(9) depending on the type of missing information. The decision variables in the optimization formulation are the box frequencies for S and R, i.e., f_{ijS} and f_{ijR}. The mean rating (μ_{iS} and μ_{iR}) and variance (σ^2_{iS} and σ^2_{iR}) of S and R are intermediate variables and j is a parameter that depends on the scale used in the survey. The formulation in Eq. (16) has a non-linear objective function with linear and non-linear constraints. The above optimization formulation needs to be solved twice, once with objective minimization for lower bound of interval and once with objective maximization for the upper bound .

Next, we discuss the estimation of intervals for Z-statistic on cumulative box frequency when the input survey data is interval in nature.

3.1.2. Estimating intervals for Z-statistic on cumulative box frequency

From the boxes used in a survey scale, i.e., $\{1,\ldots,J\}$, let J' represent the subset of the boxes for which a cumulative frequency is desired, e.g., $J' = \{J-1, J\}$ would indicate top-2-box. Let $f_{ij'S}$ represent the cumulative frequency of J' for the ith voice on S. Note that $f_{ij'S} = \sum_{j\in J'} f_{ijS}$. For the ith voice, the Z-statistic on cumulative box frequency of S with respect to R, $Z_{ij'}$, is then calculated using Eq. (17).

$$Z_{ij'} = \frac{(f_{ij'S} - f_{ij'R})}{\sqrt{\frac{f_{ij'S}(1-f_{ij'S})}{N_{iS}} + \frac{f_{ij'R}(1-f_{ij'R})}{N_{iR}}}} \tag{17}$$

Eq. (17) is similar to Eq. (14), except that the variance of the cumulative box frequency is calculated differently. If the response to a survey question is treated as a random variable, then the response would be either in the cumulative boxes we are interested or outside (including no response). So one can treat cumulative box frequency as a Binomial variable and the variance of a Binomial variable is given by $p\cdot(1-p)$, where p is the probability of occurrence for one of the events of the Binomial variable. In Eq. (17), the terms in the denominator represent the variance calculated using the Binomial distribution. With the above definition for Z-statistic on cumulative box frequency, we propose the following non-linear programming model to obtain its interval:

$$\text{Minimize and Maximize}\quad Z_{ij'} = \frac{(f_{ij'S} - f_{ij'R})}{\sqrt{\frac{f_{ij'S}(1-f_{ij'S})}{N_{iS}} + \frac{f_{ij'R}(1-f_{ij'R})}{N_{iR}}}} \tag{18a}$$

$$\sum_{j=1}^{J} f_{ijS} + f_{i0S} = 1 \text{ and } \sum_{j=1}^{J} f_{ijR} + f_{i0R} = 1 \tag{18b}$$

$$\text{Subject to}:\quad f_{ij'S} = \sum_{j\in J'} f_{ijS};\quad f_{ij'R} = \sum_{j\in J'} f_{ijR} \tag{18c}$$

$$f^{min}_{ijS} \le f_{ijS} \le f^{max}_{ijS} \text{ and } f^{min}_{ijR} \le f_{ijR} \le f^{max}_{ijR};\quad \forall j \in \{0,1,\ldots,J\} \tag{18d}$$

Eq. (18) is similar to Eq. (16) in construction, with few details changed. In Eq. (18), objective function is the Z-statistic on cumulative box frequency as defined in Eq. (17). Eq. (18b) ensures that the sum of frequencies for S and R add up to one. Eq. (18c) constrains that the cumulative box frequencies for S and R are equal to the sum of frequencies of their constituent boxes. Eq. (18d) restricts the allowed values for the frequencies of S and R to the intervals obtained using Eqs. (1)–(9) depending on the type of missing information. The decision variables in the above optimization formulation are the box frequencies for S and R, i.e., f_{ijS} and f_{ijR}. This formulation has a non-linear objective function and linear constraints. The above optimization formulation needs to be solved twice, once with objective minimization for lower bound of interval and once with objective maximization for the upper bound.

Recall that in this work, for voices with missing data, we estimate the intervals of frequencies of boxes and not the intervals for respondent level data. Use of interval respondent data would significantly increase the complexity (to the point of being intractable) of the optimization problems discussed above. Our approach makes the problem more tractable and the optimization formulations can be solved using standard software like Microsoft Excel® or MATLAB™. In fact, for our case study, we have solved Eqs. (16) and (18) using Microsoft Excel® and MATLAB™ and obtained identical results.

After converting the estimated intervals for missing data into the corresponding criteria intervals, the next step is to transform the data onto the common scale. For this, problem specific transformation functions are needed and discussed for our case study in the following section.

3.2. Transformation functions for mapping criteria values to the common scale

Herein we discuss the transformation functions for the criteria used in our case study (recall Fig. 1). Similar rules can be followed

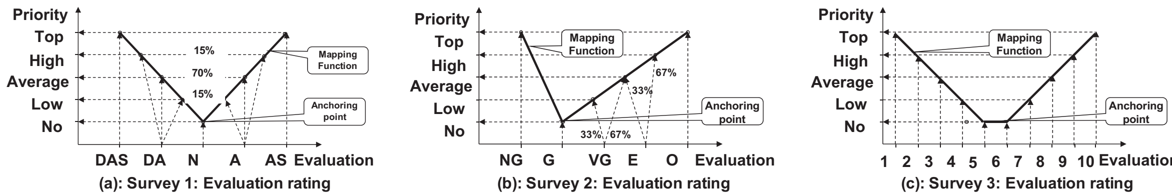

Fig. 5. Transformation functions for evaluation ratings of all three surveys.

for generating transformation functions for any other kind of criteria. Recall that for our case study, we use a 5-piont priority scale given by: "No priority, Low Priority, Average Priority, High Priority, Top Priority", as the common scale.

From Fig. 1, note that the evaluation rating of the manufacturer's vehicle is used as a criterion in all three surveys. As the three surveys use three different scales, a different transformation function for each is warranted. In addition, Survey 1 uses importance rating which needs a transformation function of its own. All three surveys use Z-statistic on mean evaluation rating and cumulative box frequencies and we propose a single transformation function for the Z-statistics.

Fig. 5 shows the transformation functions for evaluation ratings of all three surveys. If the evaluation ratings are good (e.g., "AS" or "A" in Survey 1) for the manufacturer's vehicle on a voice, those ratings should be maintained in order to remain successful in the market place. On the other hand if the evaluation ratings are poor (e.g., "DAS" or "DA" in Survey 1), those voices need to be improved in order to be competitive. This reasoning is used in proposing the transformation function shown in Fig. 5. In Fig. 5a, "A" and "DA" are mapped 15% to Low, 70% to Average, and 15% to High priority. Respondents of a survey interpret the scale in different fashions and the uncertainty in the interpretation can be captured using a distributed mapping as shown in Fig. 5a. This example shows the flexibility of our approach in incorporating various possible uncertainties. Figs. 5b and c respectively show the transformation functions for evaluation ratings of Survey 2 and Survey 3. The transformation function for evaluation rating of Survey 2 is not symmetric as the scale used in that survey is not symmetric. On the other hand, Survey 3 uses a symmetric 10-point numeric scale with 5 & 6 as anchors. Note that in Fig. 5b, "VG" is mapped 33% to Low priority and 67% to Average priority and "E" is mapped 33% to Average priority and 67% to High priority.

Fig. 6 shows the transformation function for the Z-statistic and the importance rating of Survey 1. The transformation function for Z-statistic is a step function as shown in Fig. 6a and the steps in the transformation function of Z-statistic correspond to various confidence levels associated with that metric. For example, an absolute value of 1.65 corresponds to 90% confidence level; an absolute value of 1.44 corresponds to 80% confidence level and so on. On the other hand, a linear function is used to transform importance ratings of Survey 1 to the common scale as shown in Fig. 6b. This mapping is straight-forward as the higher the importance of a voice in customers' purchase decision, the higher its priority should be.

The transformation functions proposed here would provide the belief degrees, $\beta_{n,l}$, on the common scale for the corresponding criteria. However, the above transformation functions need a precise criteria value for calculating the corresponding belief degree. For voices with missing information, as discussed in earlier sections, one would obtain interval value for the criteria of interest. These intervals needs to be transformed to the belief structure on common scale using a combination of transformation functions proposed here and the modeling techniques discussed in Section 2.4.

4. Case study from automotive domain: Results and analysis

In this section, we apply the approach developed in Section 2 to our case study making use of the new nonlinear optimization models and specific transformation functions proposed in Section 3. In the criteria hierarchy shown in Fig. 1, Survey 1 is assumed to have a weight of 0.5. Survey 2 and Survey 3 are assumed to have an equal weight of 0.25. Criteria and sub-criteria within a survey are assumed to have equal weights. Note that, these weights are assumed for demonstrative purposes only. For this paper it is not important how these weights are obtained and we suggest readers refer to (Olson, 1996; Yang, 2001) on methods for obtaining criteria weights and conducting sensitivity analysis.

As shown in Table 1, the eight voices that have to be rank ordered in our case study have different types of missing and incomplete information. Using Eqs. (1)–(9), we estimate the intervals of box frequencies from various surveys for these voices. Table 2 shows the frequency interval estimates from Survey 1 for the manufacturer vehicle's evaluation ratings, competition's evaluation ratings and importance ratings. In Table 2 and in other tables that follow, for a voice, if the row corresponding to "Max" is empty in any of the columns, it means that that voice has precise information in that column and that information is shown in the corresponding "Min" row. For Survey 1, recall from Table 1 that V3B and V3C have direct complete information and so their frequencies are estimated using Eq. (2). V5B08 has direct incomplete information and its frequencies are estimated using Eq. (1). V6A05, V6A06,

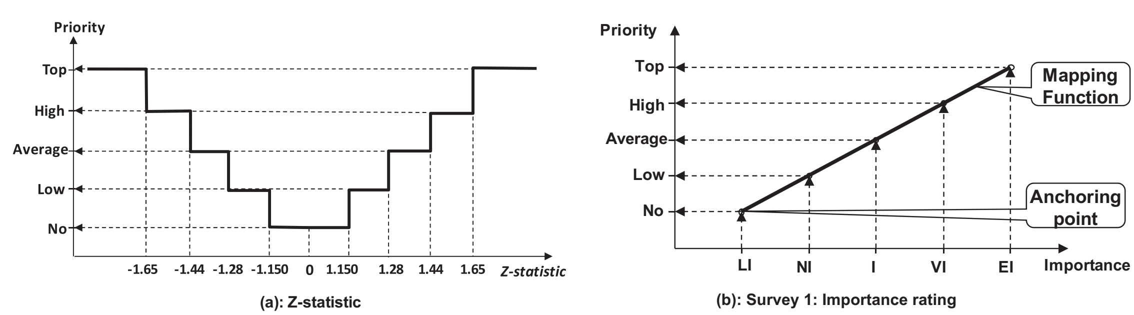

Fig. 6. Transformation function for the Z-statistic and the importance rating of Survey 1.

Table 2
Interval frequency estimates for all eight voices using available data from Survey 1.

Voice		Evaluation ratings for manufacturer's vehicle						Evaluation ratings for competition						Importance ratings					
		Incomplete	Box 1	Box 2	Box 3	Box 4	Box 5	Incomplete	Box 1	Box 2	Box 3	Box 4	Box 5	Incomplete	Box 1	Box 2	Box 3	Box 4	Box 5
V3B	Min	0.00	0.1	0.1	0.14	0.31	0.52	0.00	0.02	0.00	0.13	0.26	0.58	0.00	0.00	0.02	0.15	0.38	0.45
	Max																		
V3C	Min	0.00	0.00	0.01	0.010	0.33	0.56	0.00	0.01	0.01	0.15	0.39	0.43	0.00	0.00	0.04	0.16	0.38	0.41
	Max																		
V5B08	Min	0.05	0.050	0.050	0.06	0.25	0.64	0.05	0.050	0.050	0.04	0.20	0.71	0.02	0.020	0.07	0.16	0.38	0.37
	Max																		
V6A05	Min	0 00	0. 01	0. 03	0. 07	0. 21	0. 68	0 00	0. 00	0. 02	0. 13	0. 35	0. 51	0. 00	0. 00	0. 08	0. 11	0. 27	0. 53
	Max																		
V6A06	Min	0.00	0.02	0.07	0.15	0.33	0.43	0.00	0.01	0.07	0.17	0.33	0.43	0.00	0.02	0.07	0.25	0.41	0.25
	Max																		
V6B05	Min	0.05	0.01	0.050	0.05	0.15	0.74	0.05	0.050	0.02	0.10	0.20	0.63	0.00	0.00	0.01	0.10	0.31	0.58
	Max																		
V6B11	Min	0.00	0.01	0.06	0.14	0.28	0.51	0.00	0.01	0.03	0.13	0.35	0.48	0.00	0.01	0.08	0.25	0.37	0.29
	Max																		
V6C09	Min	0.05	0.00	0.01	0.03	0.16	0.74	0.05	0.00	0.00	0.04	0.26	0.64	0.00	0.01	0.04	0.14	0.28	0.50
	Max	0.05	0.00	0.01	0.04	0.16	0.75	0.05	0.00	0.00	0.05	0.27	0.66	0.00	0.01	0.04	0.15	0.29	0.53

Table 3
Intervals for Z-statistics on mean and cumulative box frequencies of evaluation ratings.

Voice		Survey 1			Survey 2			Survey 3		
		Z-statistic on mean evaluation rating	Z-statistic on top-box of evaluation rating	Z-statistic on top-2-box of evaluation rating	Z-statistic on mean evaluation rating	Z-statistic on top-box of evaluation rating	Z-statistic on top-2-box of evaluation rating	Z-statistic on mean evaluation rating	Z-statistic on top-2-box of evaluation rating	Z-statistic on top-4-box of evaluation rating
V3B	Min	−0.44	−0.78	−0.21	3.35	3.65	0.79	3.08	2.13	1.60
	Max									
V3C	Min	1.88	1.71	1.33	*−Inf*	*−Inf*	*−Inf*	3.14	2.36	1.18
	Max				*Inf*	*Inf*	*Inf*			
V5B08	Min	−1.03	−0.99	−0.53	80.46	0.54	−0.64	0.75	−0.11	−0.25
	Max									
V6A05	Min	1.34	2.23	0.70	2.36	2.10	1.63	−40.97	*−Inf*	−41.27
	Max							53.54	*Inf*	*Inf*
V6A06	Min	−0.04	0.00	0.09	2.94	2.35	2.65	−40.97	*−Inf*	−41.27
	Max							53.54	*Inf*	*Inf*
V6B05	Min	1.41	1.50	1.04	1.91	2.22	0.28	2.64	2.54	0.84
	Max									
V6B11	Min	−0.17	0.46	−0.52	0.85	0.73	0.39	1.16	1.63	−0.09
	Max				1.60	1.28	0.96			
V6C09	Min	0.59	1.18	−0.36	1.12	2.24	0.29	−2.24	−0.03	−1.22
	Max	1.05	1.51	0.09						

Table 4
Transformation of evaluation ratings on manufacturer's vehicle for V6C09 from Survey 1.

Common scale	Minimum belief degree	Maximum belief degree
Incomplete	0.0476	0.0476 (=min{1,0.0476})
No priority	0.0317	0.0397 (=min{1,0.0397})
Low priority	0.0256 (=0.15 × 0.0119 + 0.15 × 0.1587)	0.0262 (=min{1,0.15 × 0.0119 + 0.15 × 0.1627})
Average priority	0.1194 (=0.70 × 0.0119 + 0.70 × 0.1587)	0.1222 (=min{1,0.70 × 0.0119 + 0.70 × 0.1627})
High priority	0.0256 (=0.15 × 0.0119 + 0.15 × 0.1587)	0.0262 (=min{1,0.15 × 0.0119 + 0.15 × 0.1627})
Top priority	0.7421 (=0.0000 + 0.7421)	0.7460 (=min{1,0.0000 + 0.7460})

and V6B11 have indirectly complete information and we use Eq. (3) to estimate their frequencies. V6BB05 has indirectly incomplete information and its frequency is estimated using Eq. (4). For estimating the frequency intervals of V6C09, we use the data from its sub-voices V6C09a, V6C09b, and V6C09c. Of these three sub-voices, V6C09b is not covered in Survey 1. So we use Eq. (5) to estimate its frequency interval as minimum and maximum of the other two sub-voices. Next using Eq. (7), with $\alpha_v = 0.333$, we estimate the frequency interval of V6C09. The frequency intervals for the eight voices from Survey 2 and Survey 3 are not shown here due to space restrictions and readers can obtain the data from the Supplementary Material.

The next stage in our approach is to estimate the intervals for criteria used in evaluating voices using the frequency intervals. From Fig. 1, we can see that evaluation ratings are used as criteria in all three surveys and importance ratings are used in Survey 1. Using Eq. (16) we estimate the intervals for Z-statistic on mean evaluation ratings from all three surveys and we use Eq. (18) for estimating the intervals for Z-statistic on cumulative box frequencies. Table 3 shows the intervals for Z-statistics as calculated using

Eqs. (16) and (18). The intervals shown in Table 3 are computed using the solver functionality in Microsoft Excel® and are cross-checked with the "fmincon" routine in MATLAB™.

Note that, the intervals for Z-statistics of V3C from Survey 2 is $[-Inf, Inf]$. From Table 1, recall that for Survey 2, V3C has Type 7 data and hence the frequency intervals are [0,1] for all boxes. Similarly the intervals for Z-statistics of V6A05 and V6A06 from Survey 3 data are very large as the frequency intervals for some of their boxes are close to [0,1]. The Z-statistics' intervals for V6C09 from Survey 1 data and for V6B11 from Survey 2 data are reasonable. This is because the corresponding frequency intervals (see Table 2 for V6C09) are not large. The available information from the corresponding surveys on both these voices belongs to Type 5 (see Section 2.1.5), in which one of the corresponding sub-voices of the given voice is not covered by the survey. In Eq. (5), we proposed using the minimum and maximum frequency values of the other relevant sub-voices as the interval for the sub-voice that is not covered. Using such a rule greatly reduces the uncertainty in the missing information. If one were to use an interval of [0,1] for the sub-voices that is not covered, the intervals for the criteria could be large, similar to the case of V3C from Survey 2 data. Of course the rule we proposed in Eq. (5) is not fool-proof. For example, it will not work if only one sub-voice is covered by a survey or if all covered sub-voices have same frequencies for all boxes. However, this rule is a starting point and improvements can easily be devised depending on the problem at hand.

The next two stages of our approach are inter-related and will be discussed simultaneously in this case study. The transformation functions discussed in Section 3.2 will be used for mapping the criteria values to the five-point priority based common scale. Mapping using the transformation functions would generate the belief structure for the lower-level criteria. Because of the interval nature of the data, care should be taken in handling the belief structure as discussed in Section 2.4.

From Figs. 5 and 6b, we can see that those transformation functions are all linear in nature. So, mapping the frequency intervals of evaluation ratings from all three surveys and importance ratings from Survey 1 on to the common scale is straight-forward. One has to map the lower bound (upper bound) of a frequency interval as the lower bound (upper bound) of the belief interval using the corresponding transformation function. For example, the frequency intervals of V6C09 on evaluation ratings of manufacturer's vehicle from Survey 1 can be transformed to corresponding belief degrees as shown in Table 4. From Fig. 5a, note that Box 1 and

Table 5
Interval belief degrees on the common scale for evaluation ratings of all surveys and importance ratings of Survey 1.

Voice		Survey1		Survey 2	Survey 3
		Evaluation ratings on manufacturer's vehicle	Importance rating	Evaluation ratings on manufacturer's vehicle	Evaluation ratings on manufacturer's vehicle
V3B	Min	{0.14,0.05,0.23,0.05,0.54,0}	{0,0.02,0.15,0.38,0.45,0}	{0.01,0.04,0.12,0.21,0.55,0.07}	{0.03,0.03,0.1,0.27,0.51,0.07}
	Max				
V3c	Min	{0.1,0.05,0.24,0.05,0.56,0}	{0,0.04,0.16,0.38,0.41,0}	{0,0,0,0,0,0}	{0.05,0.05,0.1,0.22,0.51,0.07}
	Max			{1,0.67,0.66,0.67,1,1}	
V5B08	Min	{0.06,0.04,0.18,0.04,0.64,0.05}	{0,0.07,0.16,0.38,0.37,0.02}	{0.04,0.12,0.15,0.18,0.42,0.09}	{0.12,0.16,0.15,0.12,0.36,0.08}
	Max				
V6A05	Min	{0.07,0.04,0.17,0.04,0.69,0}	{0,0.08,0.11,0.27,0.53,0}	{0.03,0.07,0.13,0.2,0.49,0.07}	{0,0,0,0,0,0}
	Max				{0.48,0.55,0.55,0.89,1,0.34}
V6A06	Min	{0.15,0.06,0.28,0.06,0.45,0}	{0.02,0.07,0.25,0.41,0.25,0}	{0.08,0.1,0.15,0.21,0.36,0.11}	{0,0,0,0,0,0}
	Max				{0.48,0.55,0.55,0.89,1,0.34}
V6B05	Min	{0.05,0.02,0.11,0.02,0.75,0.05}	{0,0.01,0.1,0.31,0.58,0}	{0.02,0.04,0.1,0.18,0.59,0.08}	{0.07,0.05,0.1,0.21,0.49,0.08}
	Max				
V6B11	Min	{0.14,0.05,0.24,0.05,0.52,0}	{0.01,0.08,0.25,0.37,0.29,0}	{0.03,0.09,0.14,0.19,0.37,0.15}	{0.1,0.11,0.1,0.21,0.4,0.1}
	Max			{0.04,0.1,0.15,0.2,0.39,0.15}	
V6C09	Min	{0.03,0.03,0.12,0.03,0.74,0.05}	{0.01,0.04,0.14,0.28,0.5,0}	{0.04,0.04,0.12,0.2,0.49,0.11}	{0.04,0.1,0.12,0.14,0.37,0.23}
	Max	{0.04,0.03,0.12,0.03,0.75,0.05}	{0.01,0.04,0.15,0.29,0.53,0}		

Table 6
Interval belief degrees on the common scale for all Z-statistics from all surveys.

Voice		Survey 1			Survey 2			Survey 3		
		Z-statistic on mean evaluation rating	Z-statistic on top-box of evaluation rating	Z-statistic on top-2-box of evaluation rating	Z-statistic on mean evaluation rating	Z-statistic on top-box of evaluation rating	Z-statistic on top-2-box of evaluation rating	Z-statistic on mean evaluation rating	Z-statistic on top-2-box of evaluation rating	Z-statistic on top-4-box of evaluation rating
V3B	Min	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{0,0,0,1,0,0}
	Max									
V3C	Min	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{0,0,1,0,0,0}	{0,0,0,0,0,0}	{0,0,0,0,0,0}	{0,0,0,0,0,0}	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{0,1,0,0,0,0}
	Max				{1,1,1,1,1,1}	{1,1,1,1,1,1}	{1,1,1,1,1,1}			
V5B08	Min	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}
	Max									
V6A05	Min	{0,0,1,0,0,0}	{0,0,0,0,1,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{0,0,0,1,0,0}	{0,0,0,0,0,0}	{0,0,0,0,0,0}	{0,0,0,0,0,0}	{0,0,0,0,0,0}
	Max							{1,1,1,1,1,1}	{1,1,1,1,1,1}	{1,1,1,1,1,1}
V6A06	Min	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{0,0,0,0,0,0}	{0,0,0,0,0,0}	{0,0,0,0,0,0}
	Max							{1,1,1,1,1,1}	{1,1,1,1,1,1}	{1,1,1,1,1,1}
V6B05	Min	{0,0,1,0,0,0}	{0,0,0,1,0,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{1,0,0,0,0,0}
	Max									
V6B11	Min	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{0,0,0,0,0,0}	{0,0,0,0,0,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{0,0,0,0,1,0}	{1,0,0,0,0,0}
	Max				{1,1,1,1,0,0}	{1,1,1,0,0,0}	{1,1,1,0,0,0}			
V6C09	Min	{1,0,0,0,0,0}	{0,0,0,0,0,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{1,0,0,0,0,0}	{0,0,0,0,1,0}	{1,0,0,0,0,0}	{1,0,0,0,0,0}
	Max	{1,0,0,0,0,0}	{0,1,1,1,0,0}	{1,0,0,0,0,0}						

Box 5 of Survey 1 evaluation rating is mapped to Top priority. Box 2 and Box 4 are mapped 15% to High priority, 70% to Average priority and 15% to Low priority. Box 5 is mapped to No priority. Using this mapping the lower bound of the frequency interval of V6C09 can be mapped to the lower bound of belief degree as shown in column 2 of Table 4. From Table 2, the lower bounds for Box 1 and Box 5 of V6C09 are 0 and 0.7421 respectively and the lower bound for Top priority becomes 0.7421 as shown in the last row of column 2 in Table 4. For average priority 70% of lower bound of Box 2 (0.119) and Box 4 (0.1587) are used to obtain a lower bound of 0.1194. The third column of Table 4 shows the upper bound of belief degrees calculated using the upper bounds of the corresponding frequencies. Note that care is taken to restrict the belief upper bound to one.

Using the process shown in Table 4, the evaluation ratings on manufacturer's vehicle are transformed to the corresponding interval belief degrees on the common scale using the transformation functions shown in Fig. 5. The importance ratings from Survey 1 are transformed using the function in Fig. 6b. Table 5 shows the intervals of belief degrees on the common scale for the evaluation ratings from all the three surveys and the importance ratings of Survey 1. Note that in Tables 5 and 6 the belief degrees are shown in {No, High, Average, High, Top, Unknown} order.

The transformation function for Z-statistic is a step function as shown in Fig. 6a. Although this is non-linear, the step function properties make the mapping relatively easy. The discussion on belief structure of interval data in Section 2.4 is handy for mapping Z-statistic to the common scale. For example, the interval for the Z-statistic of top-box for V6C09 using Survey 1 data is [1.18, 1.51] (see Column 4 and last row in Table 3). This interval spans three priority grades (i.e., Low, Average, and High) as per the transformation function in Fig. 6a. These three common scale grades should be assessed to with equal likelihood. Because the transformation function in Fig. 6a is a step function, however, only one of the three grades will be assessed to for sure at any time. This means that three 0–1 binary variables need to be introduced to handle the interval data as discussed in Eq. (10). Other Z-statistics interval data can be handled in a similar way and the resultant belief degrees for all the Z-statistics from all the three surveys are shown in Table 6.

In Table 6, for the voices in the highlighted cells, the Z-statistic intervals spans across more than one grade. For those voices, binary variables as discussed in Section 2.4 should be used for guaranteeing that only one of the spanned grades are assessed to for sure at any time. These binary variables should be used during optimization for solving the interval of overall belief degree and utility. Calculating the intervals of overall belief degree for all grades of the common scale and the interval of overall utility is the final stage of our approach. The interval on overall utility is used for calculating the rank order of voices.

We solve Eq. (11) for calculating the overall belief intervals of a voice and Eqs. (12) and (13) for calculating the maximum and minimum overall utility respectively. For mapping common scale grades to utility, we use a linear utility function with zero utility for "No priority" and one for "Top priority" and 0.25 increments for the in between grades. From Fig. 1, note that there are thirteen lower level criteria for assessing a voice and each criterion has six belief degrees (including unknown/incomplete). So there are a total of seventy eight variables in each optimization run. Some of these variables, e.g., Z-statistic, could be 0–1 binary variables depending upon their intervals spanning more than one grade. So the optimization problem in Eqs. (11)-(13) is a mixed integer non linear programming problem. We have solved the optimization problems using Microsoft Excel®. Our intent here is to demonstrate the approach instead of obtaining exact solutions.

Table 7 shows the intervals for overall belief degree (Column 3), intervals for overall utility (Column 4), possible rank order of voices (Column 5), average utility value (Column 6), and ranking based on average utility value (Column 7) for all the voices in the case study. Note that average utility is the mean of min and max utilities. From the aggregated evidence, the following rankings can be generated with confidence:

1st Priority Voices:	V6B05, V3C, V6A05
2nd Priority Voices:	V3C, V6B05, V6A05, V3B, V6A06
3rd Priority Voices:	V6A05, V6B05, V3C, V3B, V6A06
4th Priority Voices:	V3B, V3C, V6A05, V6A06
5th Priority Voices:	V6C09, V6A05, V3B, V6A06
6th Priority Voices:	V6A06, V6C09, V6B11
7th Priority Voices:	V6B11, V6A06
8th Priority Voices:	V5B08

From the above evidence, one can conclude with confidence that V5B08 should be given the least priority among the eight voices assessed. From Table 7, note that V3C, V6A05, and V6A06 have longer utility intervals (i.e., difference between maximum and minimum utility). These voices have maximum uncertainty in the missing information as can be seen from Tables 5 and 6 and the same is reflected in their utilities. If the three voices V3C, V6A05 and V6A06 are temporally excluded, the following partial ranking for the

Table 7
Overall belief degree intervals, utility intervals and possible rankings for voices.

Voice		Overall belief degrees {No, Low, Average, High, Top, Unknown}	Utility interval	Possible rank	Average utility value	Ranking based on average utility
V3B	Min	{0.23, 0.02, 0.08, 0.14, 0.52, 0.01}	0.674	[2,5]	0.678	4
	Max		0 683			
V3C	Min	{0.02, 0.04, 0.11, 0.08, 0.51, 0}	0.669	[1,4]	0 776	2
	Max	{0.21, 0.22, 0.31, 0.18, 0.76, 0.15}	0.883			
V5B08	Min	{0.48, 0.04, 0.08, 0.1, 0.27, 0.02}	0.398	8	0 41	8
	Max		0.423			
V6A05	Min	{0.06, 0.02, 0.1, 0.09, 0.48, 0}	0.634	[1,5]	0 744	3
	Max	{0.23, 0.18, 0.29, 0.3, 0.73, 0.09}	0.854			
V6A06	Min	{0.21, 0.03, 0.11, 0.1, 0.29, 0.03}	0.434	[2,7]	0 561	6
	Max	{0.43, 0.12, 0.29, 0.31, 0.55, 0.1}	0.687			
V6B05	Min	{0.12, 0.01, 0.1, 0.14, 0.61, 0.02}	0.769	[1,3]	0.778	1
	Max		0 786			
V6B11	Min	{0.31, 0.07, 0.11, 0.15, 0.23, 0.02}	0.420	[6,7]	0 46	7
	Max	{0.41, 0.16, 0.2, 0.2, 0.24, 0.02}	0.500			
V6C09	Min	{0.26, 0.05, 0.07, 0.08, 0.44, 0.03}	0.569	F5 61	0 602	5
	Max	{0.26, 0.11, 0.12, 0.14, 0.46, 0.03}	0.635			

Table 8
Overall utility and ranking by ignoring missing data for voices.

Voice	Covered in Survey 1 Yes/No)	Covered in Survey 2 (Yes/No)	Covered in Survey 3 (Yes/No)	Minimum utility	Maximum utility	Possible ranking	Average utility	Ranking based on average utility
V3B	Y	Y	Y	0.605	0.642	5	0.623	5
V3C	Y	N	Y	0.832	0.837	3,2	0.834	3
V5B08	Y	Y	Y	0.385	0.410	8	0.398	8
V6A05	N	Y	N	0.881	0.893	1	0.887	1
V6A06	N	Y	N	0.822	0.864	2,3	0.843	2
V6B05	N	N	Y	0.731	0.764	4	0.747	4
V6B11	N	N	Y	0.450	0.493	7	0.471	7
V6C09	N	Y	Y	0.524	0.556	6	0.540	6

remaining five voices is fairly conclusive, even though there is missing and incomplete information associated with these five voices

1st Priority Voice:	V6B05
2nd Priority Voice:	V3B
3rd Priority Voice:	V6C09
4th Priority Voice:	V6B11
5th Priority Voice:	V5B08

Due to the missing information for voices V3C in Survey 2, V6A05 and V6A06 in Survey 3, their ranking is far from certain. For instance, V3C could be ranked to any of the 1st to 4th priority voice among the 8 voices; V6A05 to any of the 1st to 5th priority voice; V6A06 to any of the 2nd to 7th priority voice. More conclusive ranking of these voices requires more precise data, the collection of which may be expensive or even impossible. If an indicative ranking of these voices is needed for supporting decision making, the average utilities of these voices could be used, which leads to the ranking in the last column of Table 7. However, using the average utilities for ranking is not conclusive or definite. It should only be used as one of many possible indicative rankings to support further discussion for prioritization of the voices.

Table 8 shows the utility intervals and ranking of these voices using only data from the surveys in which they are actually covered. Missing information is not estimated and the common practice of ignoring the branch of criteria hierarchy (see Fig. 1) if data from a survey is absent is followed. The ER approach is used directly with the previously assumed relative weights of criteria. The results in Table 8 have a range for utilities because of incomplete responses in survey data. Comparing the results in Table 7 with those in Table 8, one can tell that the rankings suggested from Table 8 provide an illusive sense of certainty about the ranking for V3C, V6A05, and V6A06 although they both do agree that the voice V5B08 should have the lowest priority, followed by V6B11.

5. Summary

In this work, we proposed and demonstrated an approach for handling missing and incomplete information in voice of customer analysis. If a voice is not covered in a survey, we proposed using the information on relevant voices that are covered in the survey for estimating the missing information. Our approach has five main stages. In each stage, we have proposed methods and algorithms for achieving the objectives of that stage. We have demonstrated the approach with a case study, which was sampled from a much larger application, and proposed new optimization formulations for estimating intervals of criteria like Z-statistic. We showed that working with interval data on frequencies of the grades used in the survey-scale makes the estimation of intervals tractable. Also, we provided guidelines on developing transformation functions for mapping data from disparate surveys onto a common scale. Overall our investigations showed that missing information can be estimated as interval data with confidence, incomplete data can be modeled using a belief structure without loss of any information, and thus robust rankings of voices can be generated using the developed optimization models and process.

Acknowledgments

The authors sincerely thank Dr. Suresh Rajagopalan of GM R&D for all the discussions and insights.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2012.01.045.

References

Besterfield, D.H., Besterfield-Michna, C., Besterfield, G.H., Besterfield-Sacre, M., 2003. Total Quality Management, 3rd edition. Prentice Hall, New Jersey.
Bordley, R.F., 1997. Filling in incomplete survey responses. Communications in Statistics: Part B Simulation and Computation 26 (4), 1631–1651.
Bordley, R.F., Kirkwood, C.W., 2004. Multiattribute preference analysis with performance targets. Operations Research 52 (6), 823–835.
Bordley, R.F., Pollock, S.M., 2009. Decision-analytic approach to reliability-based design optimization. Operations Research 57 (5), 1262–1270.
Chan, L.K., Wu, M.L., 2002. Quality function deployment: A literature review. European Journal of Operational Research 143 (3), 463–497.
Chin, K.S., Yang, J.B., Lam, J., Guo, M., 2009. An evidential reasoning-interval based method for new product design assessment. IEEE Transactions on Engineering Management 56 (1), 142–156.
Ferson, S., Ginzburg, L., Kreinovich, V., Longpre, L., Aviles, M., 2002. Exact bounds on sample variance of interval data. ACM SIGACT News 33 (2), 108–118.
Flyer, P., Hirman, J., 2009. Missing data in confirmatory clinic trial. Journal of Biopharmaceutical Statistics 19 (6), 969–979.
Han, J., Kamber, M., 2006. Data mining: Concepts and Techniques, 2nd Edition. Elsevier, San Fransico.
Hazen, G.B., 1986. Partial information, dominance and potential optimality in multiattribute utility theory. Operations Research 34, 296–310.
James, L.C., Craig, W.K., 1991. Decision analysis applications in the operations research literature, 1970–1989. Operations Research 39 (2), 206–219.
Lee, K.S., Park, K.S., Eum, Y.S., Park, K., 2001. Extended methods for identifying dominance and potential optimality in multi-criteria analysis with imprecise information. European Journal of Operational Research 134, 557–563.
Liesio, J., Salo, A., 2012. Scenario-base portfolio selection of investment projects with incomplete probability and utility information. European Journal of Operational Research 217, 162–172.
Maddulapalli, A.K., Azarm, S., 2006. Product design selection with preference and attribute variability for an implicit value function. Journal of Mechanical Design 128 (5), 1027–1037.
Manski, C.F., 2005. Partial identification with missing data: Concepts and findings. International Journal of Approximate Reasoning 39, 151–165.
Okuda, G.E., Tauhid, S., 2008. Concept selection methods – A literature review from 1980 to 2008. International Journal of Design Engineering 1 (3), 243–276.
Olson, D.L., 1996. Decision Aids for Selection Problems. Springer-Verlag, New York.
Park, K.S., Jeong, I., 2011. How to treat strict preference information in multicriteria decision analysis. Journal of Operational Research Society 62, 1771–1783.
Sundgren, D., Danielson, M., Ekenderg, L., 2009. Warp effects on calculating interval probabilities. International Journal of Approximate Reasoning 50, 1360–1368.
Wallace, S.W., 2000. Decision making under uncertainty: Is sensitivity analysis of any use? Operations Research 48 (1), 20–25.

Wang, Y.M., Yang, J.B., Xu, D.L., Chin, K.S., 2006a. The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees. European Journal of Operational Research 175 (1), 35–66.
Wang, Y.M., Yang, J.B., Xu, D.L., 2006b. Environmental impact assessment using the evidential reasoning approach. European Journal of Operational Research 174 (3), 1885–1913.
Xie, X., Yang, J.B., Xu, D.L., Maddulapalli, A.K., Yang, Q., 2010. Uncertainty and preference modelling for multiple criteria vehicle evaluation. International Journal of Computational Intelligence Systems 3 (6), 688–708.
Xu, D.L., Yang, J.B., Wang, Y.M., 2006. The ER approach for multi-attribute decision analysis under interval uncertainties. European Journal of Operational Research 174 (3), 1914–1943.
Yang, J.B., Singh, M.G., 1994. An evidential reasoning approach for multiple attribute decision making with uncertainty. IEEE Transactions on Systems, Man, and Cybernetics 24 (1), 1–18.
Yang, J.B., Sen, P., 1994. A general multi-level evaluation process for hybrid MADM with uncertainty. IEEE Transactions on Systems, Man, and Cybernetics 24 (10), 1458–1473.
Yang, J.B., Xu, D.L. (1998). Knowledge based executive car evaluation using the evidential reasoning approach", in Advances in Manufacturing Technology XII (Baines, Taleb-Bendiab and Zhao eds), Professional Engineering Publishing, London, pp. 741–749, ISBN 1-86058-172-2.
Yang, J.B., 2001. Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty. European Journal of Operational Research 131 (1), 31–61.
Yang, J.B., Xu, D.L., 2002a. On the evidential reasoning algorithm for multiattribute decision analysis under uncertainty. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 32 (3), 289–304.
Yang, J.B., Xu, D.L., 2002b. Nonlinear information aggregation via evidential reasoning in multiattribute decision analysis under uncertainty. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 32 (3), 376–393.
Yang, J.B., Wang, Y.M., Xu, D.L., Chin, K.S., 2006. The evidential reasoning approach for MCDA under both probabilistic and fuzzy uncertainties. European Journal of Operational Research 171 (1), 309–343.
Yang, J.B., Xu, D.L., Xie, X., Maddulapalli, A.K., 2011. Multicriteria evidential reasoning decision modelling and analysis – Prioritising voices of customer. Journal of Operational Research Society 62, 1638–1654.
Yen, J., 1990. Generalizing the Dempster–Shafer theory to fuzzy sets. IEEE Transactions on Systems, Man and Cybernetics 20, 559–570.
Ying, Y., Fienberg, F., Wedel, M., 2006. Leveraging missing ratings to improve online recommendation systems. Journal of Marketing Research 43, 355–365.
Youk, A.O., Stone, R.A., Marsh, G.M., 2004. A method for imputing missing data in longitudinal studies. Annals of Epidemiology 14 (5), 354–361.