Market Segmentation Based on Risk of Misinforming Reduction

Dimitar Christozov
American University in Bulgaria, Blagoevgrad, Bulgaria

dgc@aubg.bg

Stefanka Chukova
Victoria University of Wellington, Wellington, New Zealand

stefanka@vuw.ac.nz

Plamen Mateev
Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

pmat@fmi.uni-sofia.bg

Abstract

Providing the appropriate information in the right format to the right group of potential customers of a particular product is the natural objective in market segmentation. This paper explores measures for quantifying the risk of misinforming and their usage for market studies, leading to an approach for identifying market clusters. The proposed segmentation is based on the "cost" of the additional information a particular customer needs for the revision of his/her purchase decision, so that his/her initially wrong purchase decision is appropriately corrected. The risk for wrong purchase decision is two-fold – to buy something that is not perfectly suitable or not to buy something, which is really suitable. Additional, adjusting information may reduce these risks. The cost of this additional information is calculated in a way to minimize the cost for adjusting decision regarding the buyer's tasks over all possible informing scenarios. Further, the cost for additional information is used to define the distance between clients. This distance serves to identify market segments, i.e., clusters of clients with similar information needs.

Key words: risk of misinforming, optimization problem, market segmentation

Introduction

In warranty studies we model and analyze two types of risks, the risk of malfunctioning of a product quantified by its reliability, and the risk of misunderstanding the features and qualities of

a product while deciding whether to buy it. The later risk is caused by the information asymmetry, which is a natural property of any communication process, see Christozov, Chukova, Mateev (2009a), we call it risk of misinforming. Due to the strong impact of the information asymmetry on the markets and specifically on the completion in the marketplace, studies related to information asymmetry attracted the attention of

several research teams. Mukhopadhyay, Yue and Zhu (2011) studied the impact of information asymmetry on market completion. Akan, Ata and Lariviere (2011) studied outsourcing and the associated contracting problem under information asymmetry. Knill, Minnick and Nejadmalayeri (2011) examined whether it is rational for analysts to post biased estimates and how information asymmetry and analyst experience factor into the decision aiming to minimize forecasting error. Information provided by the Seller informs different clients differently, because of the different information asymmetry. Some are informed correctly regarding all properties of the product, but the same information may mislead others. Market segmentation allows sellers to approach different categories of potential buyers differently, so that they can address them with advertising messages, or messages with the product description, in the most appropriate and beneficial way. In this paper we propose an approach for segmentation of the population of potential customers based on how they understand the seller's message and the level of needed additional adjusting information for reducing their risk of making a wrong purchase decision, i.e., the risk of misinforming.

It is natural to use the quantitative measures of the risk of misinforming for market segmentation. The earlier studies addressed mainly the warranty aspect of proposed measures for quantifying the risk of misinforming (see Christozov, Chukova, Mateev 2009b). The current study, influenced also by the ideas behind maximizing the "information gain" measure (see for example Han, Kamber, Pei (2011), Chapter 8), is addresses improvement of communication process. The idea is that maximization of information gain minimize the risk of misinforming. In this study we define a distance between customers of a given product based on the information needed to minimize the risk of misinforming and the seller's losses caused by this misinforming.

Nowadays, the market segmentation is based on the information needs of the targeted group, so ability to make a conscious and realistic purchase decision is critical. A wrong purchase decision has a negative effect on the customer, and can lead to a significant losses incurred by the seller. These losses could be of two types – based on a missed sale or due to creating an unsatisfied customer. If the product is covered by a warranty of misinforming, an unsatisfied customer could cause direct financial losses by making a legitimate warranty claim. This concept was introduced in Christozov, Chukova, Mateev (2006) where "positive" and "negative" wrong purchase decisions were identified, i.e., the "positive" wrong purchase decision will generate a legitimate warranty claim, whereas the "negative" wrong purchase decision will lead to a missed sale. Both risks are caused by the misunderstanding or misinterpretation of the message provided by the seller to the customers.

A natural question aiming to address this misunderstanding is: what must be the content of the additional message provided by seller to the customer so that his/her understanding about the real properties of the product is appropriately adjusted. Also, what is the seller's cost for providing such a message. It is clear that a particular customer must be provided with specific additional information, which reflects his/her specific product needs n_{ij} (see Christozov, Chukova, Mateev (2006)). Also, in Christozov, Chukova, Mateev (2006), was introduced a parameter q_{ij}, which reflects the minimal product requirements of the jth buyer for his/her i^{th} task. It is obvious that if a given customer needs to solve multiple tasks with the product, the message to adjust his/her decision must address the needs and minimal requirements for every specific task. In other words, for a given customer, the seller has to provide additional message addressing every task, and in this way the cost of this additional message is composed by the costs for adjusting the message for every distinct task. In this way a customer can be identified by a vector of costs for the message that will adjust his/her understanding regarding the product. Two customers, who need the same or similar adjusting message, can be considered as belonging to one and the same "market niche". Therefore, the additional information needed to ensure a correct purchase decision can be used to

define a distance between customers, and further to use this distance for clustering the customers' population or for market segmentation.

The paper is organized as follows. The next section introduces the problem's background by describing notations and model components, which are used further in the paper and provides useful references. In the following section we formulate an optimisation problem so that the Information Adjusting Cost is introduced. The third section is dedicated to the definition of distance between customers as a function of the Information Adjusting Cost for different categories of tasks. The last section provides a discussion on how to use the distance to structure the market by distinguishing clusters of clients with similar information needs. A summary of presented models and ideas and future research directions are given in the conclusion.

Background

The market is represented by the population of potential buyers of a particular product D. Let us denote this population by $B = \{b_j\}$, $j=1, 2, ..., n$, where b_j represents the jth buyer. A buyer b_j is considered a member of the market, because he/she has a set of tasks that needs to be solved for with product D. Let us also assume that $A_j = \{a_{ij}\}$, $i = 1, 2, ..., k_j$, is the set of k_j tasks that b_j needs to solve for with product D. Moreover, two additional parameters, degree of acceptance q_{ij} and needs n_{ij}, characterise the relation of b_j to his/her tasks in A_j. The degree of acceptance of the product D, q_{ij}, is a measure of the buyer's judgment regarding the minimal quality the product D has to have in order to be suitable for the buyer b_j to solve the task a_{ij}. It is described in terms of quality of the product and its value can be normalized to be within [0, 1] (see Christozov, Chukova, Mateev, 2008), where zero refers to the case where every product is acceptable and one means that only the top quality products could be satisfactory to the buyer. Here we assume that q_{ij} is known and normalized.

Furthermore, we assume that every individual task a_{ij} of buyer b_j belongs to a class of tasks A_i and that $A = \bigcup_{i=1}^{k} A_i$ is the set of all categories of tasks, which are common for many, not necessarily all, buyers in B. We assume that every b_j has tasks from each of the classes A_i, but for some of these tasks his/her need n_{ij} is zero. The product D is produced in a way to allow solving the tasks from A. It has the potential or capability to solve task from a given class with probability p_i. In the case that a buyer needs to solve for a task that is not among the tasks the product D is designed for, corresponding $p_i = 0$. The need of the b_j to solve for a task from category A_i is n_{ij}, where $0 \leq n_{ij} \leq 1$. If $n_{ij} = 0$, then b_j does not need to solve for any task from A_i, whereas, if $n_{ij}=1$, b_j definitely needs to solve for a task from A_i. In case of $n_{ij} = 0$, then $q_{ij}=0$, i.e., if b_j does not need to solve for a task from A_i, then any product is acceptable for him/her. The value of n_{ij} is completely subjective and could be estimated using customers' surveys.

The seller advertises the product D by sending a message to B describing the properties of D. Based on this message b_j assesses the probability $\hat{p}_{ij} = \hat{p}(a_{ij})$ that the product D is suitable to solve for his/her task a_{ij}. Following this assessment, b_j makes his/her purchase decision based on the comparison between $\{\hat{p}_{ij}\}$ and $\{q_{ij}\}$ over all tasks from his/her set A_j, the values of p_i and $\hat{p}_{ij}$ may differ significantly. There are six cases, described in Christozov, Chukova, Mateev (2006), addressing the risks r_{ij} associated with such a purchase decision:

1. $p_i < \hat{p}_{ij} < q_{ij}$ - the product is not suitable to solve for task a_{ij}, the buyer's estimation of the suitability of the product is **optimistic,** i.e., $p_i < \hat{p}_{ij}$, and **below the degree of acceptance,** thus the decision is **negative** and **correct** and $r_{ij}=0$;

2. $p_i < q_{ij} < \hat{p}_{ij}$ - the product is not suitable to solve for task a_{ij}, the buyer's estimation of the suitability of the product is **optimistic** and **above the threshold of acceptance,** thus the decision is **positive** and **wrong,** and $r_{ij}=1$;

3. $q_{ij} < p_i < \hat{p}_{ij}$ - the product is suitable to solve for task a_{ij}, the buyer's estimation of the suitability of the product is **optimistic** and **above the threshold of acceptance** , thus the decision is **positive** and **correct,** and $r_{ij}=0$;

4. $\hat{p}_{ij} < p_i < q_{ij}$ - the product is not suitable to solve for task a_{ij}, the buyer's estimation of the suitability of the product is **pessimistic,** i.e., $\hat{p}_{ij}$ is less than p_i, and **below the threshold of acceptance,** thus the decision is again **negative** and **correct,** and $r_{ij}=0$;

5. $\hat{p}_{ij} < q_{ij} < p_i$ - the product is suitable to solve for task a_{ij}, the buyer's estimation of the suitability of the product is **pessimistic** and **below the threshold of acceptance**, thus the decision is **negative** and **wrong,** and $r_{ij}=1$;

6. $q_{ij} < \hat{p}_{ij} < p_i$ - the product is suitable to solve for task a_{ij}, the buyer's estimation of the suitability of the product is **pessimistic** and **above the threshold of acceptance**, thus the decision is **positive** and **correct,** and $r_{ij}=0$.

The two wrong decisions relate to the two categories of risks for the seller: case 2 ($p_i < q_{ij} < \hat{p}_{ij}$) corresponds to missed sales and missed profit; and case 5 ($\hat{p}_{ij} < q_{ij} < p_i$) corresponds to the potential dissatisfaction of the buyer. If the product is covered by a misinforming warranty contract (see Christozov, Chukova Mateev, 2009a), the second risk could result in direct financial losses due to customer's making a legal warranty claim. These risks could be non-zero only if q_{ij} is between p_i and $\hat{p}_{ij}$. It is natural to expect that the seller's aim is to reduce both of these risks. To do that the seller has to develop information policy (advertise or distribute additional messages) to address buyers' specific information needs so that the gap between p_{ij} and $\hat{p}_{ij}$ is reduced. This will reduce the risk for experiencing either of the two types of losses. This requires identifying the clusters of buyers with similar understanding of the original message and need of similar additional information to adjust their purchase decision. In our modelling, onwards, we assume that providing such additional information comes with certain cost to the seller.

Optimization Problem:
Minimizing Information Adjusting Cost

In what follows we aim to formulate an optimisation problem related to the risk of misinforming, so that the related "costs" are minimized. We introduce an appropriate cost measure and study the case of limited resources.

As mentioned above, there are two types of risks:

$p_i < q_{ij} < \hat{p}_{ij}$ - the product is not suitable to solve for task a_{ij}, the buyer's estimation of the suitability of the product is **optimistic** and **above the threshold of acceptance,** thus the decision is **positive** and **wrong,** and $r_{ij}=1$. This situation creates an opportunity for a warranty claim. The buyer will realise that the product is not suitable to address his/her expectations regarding task a_{ij} and would be prompted to make a warranty claim against the product. It could create a number of

dissatisfied customers and affect the overall market reputation of the seller. Denote this type of risk by r_{ij}^{wc}.

$\hat{p}_{ij} < q_{ij} < p_i$ - the product is suitable to solve for task a_{ij}, the buyer's estimation of the suitability of the product is **pessimistic** and **below the threshold of acceptance**, thus the decision is **negative** and **wrong,** and $r_{ij}=1$. This situation creates an opportunity for a missed sale, which leads to a reduced profit for the seller. Denote this type of risk by r_{ij}^{ms}.

It is easy to notice that the risk of misinforming is non-zero only if $q_{ij} \in (min\,(p_i, \hat{p}_{ij})$, $max\,(p_i, \hat{p}_{ij})$), i.e., q_{ij} is between p_i and $\hat{p}_{ij}$. Let the preference of the seller in minimising these two types of risks $(r_{ij}^{ms}, r_{ij}^{wc})$ is given by $(\tau,\ 1\text{-}\ \tau)$, where $0 \le \tau \le 1$. Therefore, the objective function of the optimisation problem related to the b_j could be formulated as follows:

$$min\ Z = \tau\ R_j^{ms} + (1\text{-}\ \tau)\ R_j^{wc},$$

where $R_j^{ms} = \sum_{i=1}^{k} r_{ij}^{ms}$ and $R_j^{wc} = \sum_{i=1}^{k} r_{ij}^{wc}$. Onwards, we will assume that the risks $(r_{ij}^{ms}, r_{ij}^{wc})$ are zero whenever $p_i < q_{ij} \le \hat{p}_{ij}$ or $\hat{p}_{ij} \le q_{ij} < p_i$, i.e., the risk is zero if the acceptance level q_{ij} is equal to the probability $\hat{p}_{ij}$. Next, let us discuss the possible constraints for this optimization problem. Assume that q_{ij} and p_i are known whereas $\hat{P}_{ij}$ is uncertain, i.e., it is a random variable with known probability distribution. We assume that $\hat{P}_{ij}$ has a standard beta distribution with parameters $(\alpha_{ij}, \beta_{ij})$, i.e., $\hat{P}_{ij} :\ Beta(\alpha_{ij}, \beta_{ij})$ and the reported (observed) value $\hat{p}_{ij}$ is the mode of this distribution. For more on the standard beta distribution with parameters $(\alpha_{ij}, \beta_{ij})$, including the probability density function, mean value μ_{ij} and variance σ_{ij} , see Appendix II. Therefore for the a_{ij} task of b_j, the triple $(\hat{p}_{ij}, q_{ij}, p_i)$ represents the level of informing induced by the sellers' message regarding the features of product D. Assume that for b_j it is true that $(r_{ij}^{ms}, r_{ij}^{wc}) \ne (0, 0)$. The seller could provide an additional information to the buyer, so that the buyer's risks $(r_{ij}^{ms}, r_{ij}^{wc}) = (0, 0)$. For providing this additional information the seller will incur certain cost, but will benefit from the reduction of the risk of misinforming of b_j with respect of task a_{ij}. Assume that the effect of the information in the additional message is measured by the shift of the initial $Beta(\alpha_{ij}, \beta_{ij})$ to $Beta(\tilde{\alpha}_{ij}, \tilde{\beta}_{ij})$, which preserves the variance of the initial distribution and assures that the mode of the new beta distribution $Beta(\tilde{\alpha}_{ij}, \tilde{\beta}_{ij})$ is equal to q_{ij}, i.e., assures that the corresponding risk of misinforming becomes zero. How bad the initial message was (measured by $Beta(\alpha_{ij}, \beta_{ij})$, Figure1) is estimated by the following "cost": the absolute value of the difference between the mean values of the initial beta distribution μ_{ij} and the mean value $\tilde{\mu}_{ij}$ of the beta distribution induced, see Figure 2, by the new message, normalised by the common standard deviation σ_{ij} of these two distributions.
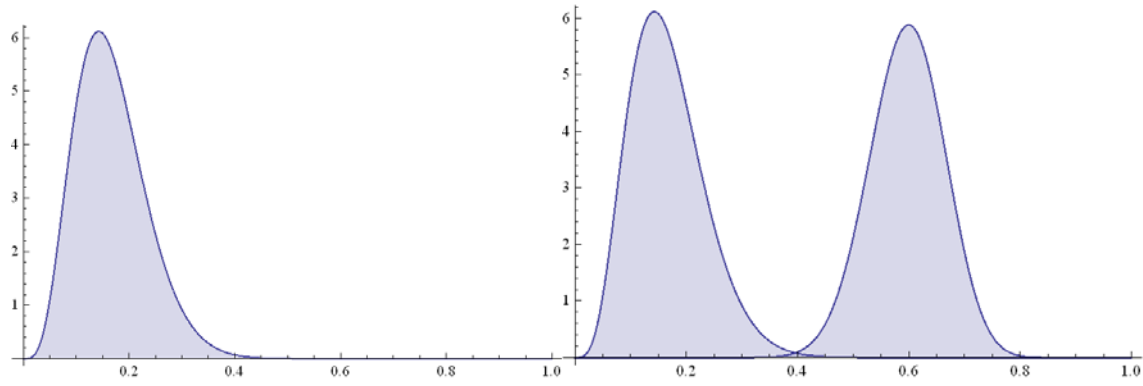
Figure 1: Initial Beta Distribution   Figure 2. Initial and induced Beta distributions

We are almost ready to formulate the optimization problem aiming to identify the "best" new message for b_j, given limited seller's resources. The last component in the formulation of the optimization problem will address the entire set of tasks of b_j, not only one of his/her tasks a_{ij}, as it was in the context above. Let v_{ij}, $0 \leq v_{ij} \leq 1$, $i = 1, ..., k_j$ be the importance, from seller's viewpoint, in addressing with the new message the misinforming related to the ith task of the b_j. Due to marketing or competitive reasons the seller may have preferences in investing his/her resources towards a specific subset of tasks. In the formulation of the optimization problem, the v_{ij} parameters are assumed to be known. Therefore, we are ready to consider the following problem:

$$\min Z = \tau R_j^{ms} + (1-\tau) R_j^{wc},$$

$$\text{subject to:} \qquad \sum_{i=1}^{k} v_{ij} \frac{\left| \mu_{ij} - \widetilde{\mu}_{ij} \right|}{\sigma_{ij}} \leq C$$

where $R_j^{ms} = \sum_{i=1}^{k} r_{ij}^{ms}$, $R_j^{wc} = \sum_{i=1}^{k} r_{ij}^{wc}$ and C is a known constant, which represents the limited resources of the seller.

What type of solution for this optimization problem are we looking for? What is the meaning of this solution? We need to identify a transformation form $Beta(\alpha_{ij}, \beta_{ij})$ to $Beta(\widetilde{\alpha}_{ij}, \widetilde{\beta}_{ij})$, so that the objective function Z is minimized and the "cost" constraint is satisfied.

Let us also denote the cost of transformation for a given task as

$$c_{ij} = v_{ij} \frac{\left| \mu_{ij} - \widetilde{\mu}_{ij} \right|}{\sigma_{ij}} . \qquad\qquad (1)$$

Next, using an example, we illustrate the ideas presented in this section.

An Example

Let us assume that the set of tasks of the jth buyer consists of six tasks: $A_j = \{a_{1j}, a_{2j}, ..., a_{6j}\}$, such that $\hat{p}_{1j} < q_{1j} < p_1$, $\hat{p}_{4j} < q_{4j} < p_4$ and $p_3 < q_{3j} < \hat{p}_{3j}$. In addition, it is known that $q_{ij} = 0.6$, $q_{3j} = 0.55$, $q_{4j} = 0.65$; and let $p_1 = 0.7$, $p_3 = 0.3$, $p_4 = 0.8$. Also it is known that the initial beta distributions are as follows: $\hat{P}_{1j} : Beta(5,25)$, $\hat{P}_{3j} : Beta(23,7)$ and $\hat{P}_{4j} : Beta(3,12)$, which leads to the following values for $\hat{p}_{1j} = 0.142857$, $\hat{p}_{3j} = 0.7857$ and $\hat{p}_{4j} = 0.153846$. For the remaining three tasks $(r_{ij}^{ms}, r_{ij}^{wc}) = (0, 0)$, $i = 2,5,6$. Therefore we have $r_{1j}^{ms} = r_{4j}^{ms} = r_{3j}^{wc} = 1$. Next, assuming particular values for the model parameters such as: $v_{1j} = 0.2$, $v_{3j} = 0.4$, $v_{4j} = 0.7$, $\alpha = 0.7$, $C = 1.5$, we will illustrate the process of solving the optimization problem formulated in the previous section:

min $Z = 0.7\, R_j^{ms} + 0.3 R_j^{wc}$,

$$\text{subject to:} \qquad \sum_{i=1,3,4} v_{ij} \frac{\left|\mu_{ij} - \widetilde{\mu}_i\right|}{\sigma_{ij}} \le \frac{3}{2}$$

What does this particular set of model parameters reveal? The value of $\alpha = 0.7$ suggests that the seller's main concern is to minimise the losses due to missed sales, i.e., the seller places higher priority on reducing the risk of the missed sales compared to the risk of receiving a warranty claim. Moreover, the values of v_{ij}, show that the seller places the highest importance in eliminating the misinforming related to task a_{4j}, whereas he is least concern with the risk associated with task a_{1j}. The value C=1.5 represents the upper bound of the resource (or the "cost") the seller is allowed or willing to invest in the new message aiming to reduce the risk of misinforming. Next we present the solution of the problem.

- For task a_{1j} it is given that $\hat{p}_{1j} \le q_{1j} < p_1 \Rightarrow 0.142857 \le 0.6 < 0.7$, i.e., the risk $r_{1j}^{ms}=1$ is for wrong negative decision. We need to find a transformation, preserving the standard deviation, of the initial $Beta(5,25)$ with mode equal to $\hat{p}_{1j} = 0.142857$ to a new $Beta(\alpha, \beta)$ with mode equal to $q_{1j} = 0.6$. Using the popular software MATHE-MATICA (see http://www.wolfram.com/mathematica/), we were able to write a script that searches for this type of transformation. For task a_{1j} the new beta distribution with the required properties is $Beta(31.4405, 21.2936)$ and the "cost" of this transformation is

$$c_{1j} = v_{1j} \frac{\left|\mu_{1j} - \widetilde{\mu}_{1j}\right|}{\sigma_{1j}} = 6.41729 \,.$$

- For task a_{3j} it is given that $p_3 < q_{3j} \le \hat{p}_{3j} \Rightarrow 0.3 < 0.55 \le 0.7857$, i.e., the risk $r_{3j}^{wc}=1$ is for wrong positive decision. We need to find a transformation, preserving the standard deviation, of the initial $Beta(23,7)$ with mode equal to $\hat{p}_{3j} = 0.7857$ to a new $Beta(\alpha, \beta)$ with mode equal to $q_{3j} = 0.55$. Again using the MATHEMATICA script, we were able to determine that for a_{3j} the new beta distribution with the required properties is $Beta(22.9616, 18.9685)$ and the "cost" of this transformation is

$$c_{3j} = v_{3j} \frac{\left| \mu_{3j} - \widetilde{\mu}_{3j} \right|}{\sigma_{3j}} = 0.28836 .$$

- At last, for task a_{4j} it is given that $\hat{p}_{4j} \leq q_{4j} < p_4 \Rightarrow 0.153846 \leq 0.65 < 0.8$, i.e., the risk $r_{4j}^{ms} = 1$ is for wrong negative decision. We need to find a transformation, preserving the standard deviation, of the initial $Beta(3,12)$ with mode equal to $\hat{p}_{4j} = 0.153846$ to a new $Beta(\alpha, \beta)$ with mode equal to $q_{4j} = 0.65$. Again using the MATHEMATICA script, we were able to determine that for a_{4j} the new beta distribution with the required properties is $Beta(14.0898, 8.04836)$ and the "cost" of this transformation is

$$c_{4j} = v_{4j} \frac{\left| \mu_{4j} - \widetilde{\mu}_{4j} \right|}{\sigma_{4j}} = 4.36449 .$$

We have evaluated the "costs" of the transformations related to each of the tasks of interest. Next, we consider all possible transformations scenarios and select the "best" one that satisfies the resource constraint and leads to the smallest value of the objective function. We present our findings in Table 1.

# scenario	# transform	Task 1	Task 3	Task 4	Cost of the Scenario	Constraints Satisfied?
1	3	yes	yes	yes	5.49404	No
2	2	yes	yes	no	2.4389	No
3	2	yes	no	yes	4.3406	No
4	2	no	yes	yes	4.20858	No
5	1	no	no	yes	3.05514	No
6	1	no	yes	no	1.15344	Yes
7	1	yes	no	no	1.28546	Yes

Table 1. Cost of all informing scenarios

From Table 1 it is easy to conclude that only the last two scenarios satisfy the constraints of the optimization problem. Next, we have to make a decision which of these two scenarios to use as a bases in forming our new message. Let us compute the value of the objective function over each of these scenarios. We have for scenario 6, $Z_6 = 0.7 \times 2 + 0.3 \times 0 = 1.4$, whereas for scenario 7, $Z_7 = 0.7 \times 1 + 0.3 \times 1 = 1$. Thus, the minimum value of the objective function is reached over scenario 7, and therefore the new message should be designed so that it targets task 1 of the jth buyer.

Definition of Distance based on Information Adjusting

In the previous section we outlined an approach for computing the optimal cost for composing an adjusting message for a single independent potential buyer. For minimizing the overall cost, it is important to identify groups of buyers, which can be targeted simultaneously with one and the same adjusting message. This is a typical cluster analysis task where the essential part is to define

distance between any two objects $dist(b_{j_1}, b_{j_2})$, belonging to the studied population (the set of all buyers *B*).

Let us assume that the risks of wrong purchase decision for buyer b_j with respect of his/her set of tasks are $r(b_j)=\{r_{1j}, r_{2j}, ..., r_{kj}\}$, which is a binary vector, composed only by zeros and ones. The buyer needs additional information, which has to address the properties of the product *D*, which are essential for the description of the performance of the product *D* regarding tasks with risk of wrong purchase decision equal to one. The initial message informed correctly the buyer regarding the properties of the product relevant to the tasks with risk of wrong purchase decision equal to zero. In other words, the additional, adjusting information should address the information needs of buyer b_j, so that the risk of wrong decision is reduced, i.e., it has to address only the properties of product *D* that have impact on its performance related to solving tasks with non-zero risk. The adjusting message has to address the buyers' different information needs and it is worth to identify market segments based on the buyers' needs of additional information. To identify different clusters of buyers, we need to define a distance between buyers according to their needs of adjusting information.

Next, we propose two possible definitions for the "distance" between two buyers:

Task-based distance – the message addresses product's properties relevant only to the risky tasks. These vectors $\{r_{1j}, r_{2j}, ..., r_{kj}\}$, $j=1,...,n$ hold asymmetric binary values. The distance can be defined by counting differences (see for example Han, Kamber, Pei (2011), section 2.4.3, page 70).

Cost-based distance – in composing the adjusting message the "cost" of the adjusting message is taken into account. This measure uses vectors $\{c_{1j}, c_{2j}, ..., c_{kj}\}$, which includes continuous values. The "costs" of the transformation needed to convert non-zero risks tasks into zero risk tasks are calculated according to the formulae (1) in the previous section. The Minkowski distance $dist(b_{j_1}, b_{j_2})$ between buyers b_{j_1} and b_{j_2} based on the corresponding cost (see Han, Kamber, Pei (2011), section 2.4.4, page 72) is given by the formulae

$$dist(b_{j_1}, b_{j_2}) = \sqrt[h]{\sum_{i=1}^{k} \left| c_{ij_1} - c_{ij_2} \right|^h} \ .$$

If in the above formulae *h = 1*, we have Manhattan distance, and if *h = 2* – Euclidean distance. Moreover, when $h \to \infty$ we get, the so called, Chebyshev's distance. It is defined as follows:

$$dist(b_{j_1}, b_{j_2}) = \lim_{h \to \infty} (\sum_{i=1}^{k} \left| c_{ij_1} - c_{ij_2} \right|^h)^{1/h} = \max_{i=1}^{k} \left| c_{ij_1} - c_{ij_2} \right| \ .$$

So, any of the above distances could be used to identify clusters of buyers. As follow up from the example in the previous section, it seems that the Chebyshev's distance is the simplest and reflects well our ideas in optimizing the effect of the adjusting message under given cost constraint. The cost-based distance will identify buyers with small maximal cost distance over all tasks. These buyers could be grouped together in a cluster that could be targeted by one and the same appropriately designed adjusting message. These clusters will form the segmentation of the market based on the risk of misinforming.

Conclusion

In this paper we presented an approach to structure the market by segmentation based on information needed by customers to make accurate risk-free purchase decision. Providing the necessary information in a correct format to the targeted group of customers is the natural objective of market segmentations. The proposed segmentation aims to reduce the risk of misinforming leading to two categories of losses – losses due to missed sales/purchases and losses due to customer dissatisfaction. These risks may have monetary implications for both parties - the sellers as well as the buyers.

Two approaches to define distance between potential buyers, based on the process of adjusting the message on the properties of the product, are presented. The first addresses only the content of the message for different groups of buyers. The second one is based on minimizing the cost of the adjusting message. Either of these approaches can be used for further clustering of the market.

There are several aspects of the proposed model that need additional extension and clarification, such as:

1. Develop a procedure to identify whether the observed risks represent outliers and are not worthy to additional investment in fine-tuning the message.

2. How and what information to collect in order to evaluate these risks?

    a. It is obvious that it is much easier to obtain information regarding the risk caused by users' dissatisfaction. If the product is covered by warranty of misinforming this information will be recorded in the corresponding warranty database.

    b. It is much more difficult to collect information regarding the risk of missed sales. To collect this information the seller has to invest in providing special surveillance or careful recording, surveying and analysis of all purchase inquiries.

3. How to assess the risk related to the changes in the needs of a particular buyer on the use of the product and some of the particular product features?

In our follow up studies we will make an attempt to address all or part of the listed above extensions of the proposed model.

References

Akan, M., Ata, B., & Lariviere, M. A. (2011). Asymmetric information and economies of scale in service contracting. *Manufacturing and Service Operations Management*, *13*(1) , 58-72.

Brayn, F. J. M. (2005). *Multivariate statistical methods - A primer* (3rd ed). Chapman and Hall.

Christozov, D., Chukova, S., & Mateev, P. (2006). A measure of risk caused by information asymmetry in e-commerce. *Journal of Issues in Informing Science and Information Technology*, *3*, 147-158.

Christozov, D., Chukova, S., & Mateev, P. (2008). Warranty and the risk of misinforming: evaluation of the degree of acceptance. *Journal of Issues in Informing Science and Information Technology*, *5*, 667-677.

Christozov, D., Chukova, S., & Mateev, P. (2009a). On two types of warranties: Warranty of malfunctioning and warranty of misinforming. *Asia-Pacific Journal on Operation research*, *26*(3), 399-420.

Christozov, D., Chukova, S., & Mateev P. (2009b). Chapter 11: Informing processes, risks, evaluation of the risk of misinforming. In T. G. Gill & E. Cohen (Eds.), *Foundations of Informing Science* (pp. 323-356). Santa Rosa, CA: Informing Science Press.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques*. Morgan Kaufmann.

Knill, A., Minnick, K., & Nejadmalayeri, A. (2011). Experience, information asymmetry, and rational forecast bias. *Review of Quantitative Finance and Accounting*, 1-32.

Mukhopadhyay, S. K., Yue, X., & Zhu, X. (2011). A Stackelberg model of pricing of complementary goods under information asymmetry. *International Journal of Production Economics*, *134*(2), 424-433.

Owen, C. B., (2008). *Parameter estimation for the beta distribution*. MSc thesis, Brigham Young University, Utah, USA

Appendix 1. Notations and definitions, (see Christozov, Chukova, Mateev (2011)

Notation	Definition
D	the product
$B = \{b_j\}, j=1, 2, ..., n$	buyers
$A_j = \{a_{ij}\}, i = 1, 2, ..., k_j$	tasks, which b_j needs to solve by using the product
$A = \bigcup_{j=1}^{n} A_j$	set of tasks of all buyers
$A_i^*, i=1, 2, ..., k$	categories of tasks
n_{ij}	the need of b_j to solve his/her task a_{ij}. $0 \le n_{ij} \le 1$
q_{ij}	degree of acceptance. The minimal quality, which the product must possess to meet the customer b_j expectations regarding his/her task a_{ij} threshold.
$p_i = p(A_i^*)$	probability that the product will solve problems from category A_i^*. Or the level to which the product D may satisfy the buyers needs regarding the tasks from this category.
$\hat{p}_{ij} = \hat{p}(a_{ij})$	subjective assessment of the buyer b_j regarding the probability (level of satisfaction) that the product will be suitable for solving his/her task a_{ij}.
r_{ij}	indicator of the decision correctness $r_{ij}=0$ if the decision is correct; $r_{ij}=1$ means wrong decision.

Appendix II. Standard beta distribution with parameters (α, β)

The Beta distribution is a continuous probability distribution defined on [0,1]. For more details on the standard beta distribution see Owen C. B., (2008). The probability density function (*pdf*) of this distribution is defined by $f_{\hat{p}}(x) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$ for x in [0,1], where

$$B(\alpha, \beta) = \int_0^1 x^\alpha (1-x)^{\beta-1} dx$$

The mean value of $\hat{P} : Beta(\alpha, \beta)$ is equal to $\mu = \dfrac{\alpha}{\alpha + \beta}$, its mode

is given by $\hat{p} = \dfrac{\alpha - 1}{(\alpha + \beta - 2)}$ and the variance by $\sigma^2 = \dfrac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$.

Moreover, it is well known that the **pdf** is unimodal and strongly skewed to the right if $\alpha > 1, \beta > 1$ and $\alpha \leq \beta$. If we keep α fixed and increase the value of β it makes the density more picked. The distribution gets less skewed and the mode approaches 0.5 as α and β approach each other. The distribution is symmetric around 0.5 if $\alpha = \beta$. If $\alpha = \beta = 1$ then the beta distribution becomes the well known uniform distribution over [0, 1]. In the case of left skewed **pdf**, if we keep β fixed and increase the value of α it makes the density more picked. The **pdf** is right skewed if the values of α and β were switched. If P is $Beta(\alpha, \beta)$, then $Q=1-P$ is $Beta(\beta, \alpha)$.

Beta distribution is very flexible because it's two positive shape parameters α and β are frequently used to estimate an unknown proportion parameter p, which is between 0 and 1. Increasing the parameter α is equivalent to evidence of positive behaviour, because the mean value and mode are moving to right. The opposite case – increasing parameter β – leads to decreasing the mean value and mode, which corresponds to negative behaviour. In both cases the variance is decreasing. So, the two parameters of the beta distribution may be used to weight all pros and cons regarding the values of the proportion parameter p.

In the table below the mean value, mode and standard deviation are shown for some combinations of the parameters α and β.

		α							
		2	3	5	8	12	20	30	
	2	0.50	0.60	0.71	0.80	0.86	0.91	0.94	mean
		0.50	0.67	0.80	0.88	0.92	0.95	0.97	mode
		0.22	0.20	0.16	0.12	0.09	0.06	0.04	stdev
	3	0.40	0.50	0.63	0.73	0.80	0.87	0.91	mean
		0.33	0.50	0.67	0.78	0.85	0.90	0.94	mode
		0.20	0.19	0.16	0.13	0.10	0.07	0.05	stdev
	5	0.29	0.38	0.50	0.62	0.71	0.80	0.86	mean
		0.20	0.33	0.50	0.64	0.73	0.83	0.88	mode
		0.16	0.16	0.15	0.13	0.11	0.08	0.06	stdev
β	8	0.20	0.27	0.38	0.50	0.60	0.71	0.79	mean
		0.13	0.22	0.36	0.50	0.61	0.73	0.81	mode
		0.12	0.13	0.13	0.12	0.11	0.08	0.07	stdev
	12	0.14	0.20	0.29	0.40	0.50	0.63	0.71	mean
		0.08	0.15	0.27	0.39	0.50	0.63	0.73	mode
		0.09	0.10	0.11	0.11	0.10	0.08	0.07	stdev
	20	0.09	0.13	0.20	0.29	0.38	0.50	0.60	mean
		0.05	0.10	0.17	0.27	0.37	0.50	0.60	mode
		0.06	0.07	0.08	0.08	0.08	0.08	0.07	stdev
	30	0.06	0.09	0.14	0.21	0.29	0.40	0.50	mean
		0.03	0.06	0.12	0.19	0.28	0.40	0.50	mode
		0.04	0.05	0.06	0.07	0.07	0.07	0.06	stdev

Biographies

Dimitar Christozov is a Professor of Computer Science at the American University in Bulgaria, Blagoevgrad 2700, Bulgaria since 1993, e-mail: dgc@aubg.bg. He has more than 30 years of experience in areas as computer science, quality management and information systems. He graduated Mathematics from Sofia University "St. Kliment Ohridski" in 1979. He completed his PhD thesis "Computer Aided Evaluation of Machine Reliability" in 1986. In 2009 he completes his D.Sc. thesis on "Quantitative Measures of Informing Quality". In ICTT "Informa" (1986-1993) Dr. Christozov was involved in establishing the national information network for technology transfer and conducted research in the areas of technologies assessment, integral quality measures and information systems for quality management. In these areas he was recognized as one of the leading experts in Bulgaria. Professor Christozov has more than 80 publications as separate volume, journal papers and papers in refereed proceedings. He is a founding member of Informing Science Institute and chair of Bulgarian Informing Science Society; and founding member of the Bulgarian Statistical Society.

Dr. Stefanka Chukova is an Associate Professor in Operations Research at the School of Mathematics, Statistics and Operation Research, Victoria University of Wellington, Private Bag 600, Wellington, New Zealand, e-mail: schukova@msor.vuw.ac.nz. She has a PhD and MSc in Mathematics (concentration in Probability and Statistics) and BSc in Mathematics from University of Sofia, Sofia, Bulgaria. Her research interests are in applied stochastic models, warranty analysis, reliability and queuing. She has more than 80 publications and has presented papers at national and international conferences.

Dr. Plamen Mateev is Associate Professor in Faculty of Mathematics and Informatics, Sofia University "St.Kliment Ohridski", Department "Probability, Operation Research, Statistics", Bulgaria, 1164 Sofia, 5, J.Boucher str., e-mail: pmat@fmi.uni-sofia.bg.

His MSc in Mathematical Statistics is from Sofia University and his PhD is from Moscow State University. The research interests are in communication theory, applied statistics, statistical software and applications. More than 70 papers are published in scientific journals and proceedings of scientific conferences. He was the Chair of Bulgarian Statistical Society and member of ENBIS and Bulgarian Informing Science Society.