# Visual Survey Analysis in Marketing

**Author, Affiliation, Country**

## ABSTRACT
In this chapter, the authors present a use and visualization of the ordinal evaluation (OrdEval) algorithm as a promising technique to study questionnaire data. The OrdEval algorithm is a general tool to analyze data with ordinal attributes, including surveys. It has many favorable features, including context sensitivity, ability to exploit meaning of ordered features and ordered response, robustness to noise and missing values in the data, and visualization capability. The authors select customer (dis)satisfaction analysis, an important problem from marketing research, as a case study and present visual analysis on two practical applications: business-to-business and costumer-to-business customer satisfaction studies. They demonstrate some interesting advantages offered by the new methodology and visualization and show how to extract and interpret new insights not available with classical analytical toolbox.

## BACKGROUND
In recent years we have observed large changes in economy in general and marketing in particular as a result of internet expansion, globalization, and ubiquitous information availability. One of the scientific fields which gained momentum as a result of this was data analysis under various names: statistics, data mining, machine learning, intelligent data analysis, knowledge discovery. Many new data analysis techniques emerged which exploit availability of more and different data from several sources, and increased computational power of nowadays computers. Some examples of these techniques are support vector machines, text analytics, association rules, ensemble techniques, subgroup discovery, etc. These techniques have been accepted into analytics' standard toolbox in many disciplines: genetics, engineering, medicine, vision, statistics, marketing, etc.

The *OrdEval* algorithm (Robnik-Šikonja & Vanhoof, 2007) is a novel analytical tool which emerged in data mining context aiming to evaluate the importance and the impact of various factors in the given data (e.g., survey). For example, in the analysis of customer satisfaction data for a particular product/service, OrdEval can determine the importance of each product's feature to the overall customer's satisfaction, and also indicate the thresholds where satisfaction with the individual feature starts having a strong positive or negative impact on the overall satisfaction. The output of OrdEval are probabilistic factors indicating the probability that increase/decrease in the individual feature or the feature's value will have impact on the dependent variable. The intuition behind this approach is to approximate the inner workings of the decision process taking place in each individual respondent, which forms the relationship between the features and the response. If such brain introspection would be possible one could observe a causal effect that the change of a feature's value has on the response value. By measuring such an effect we could reason about the importance of the feature's values and the type of the attribute. Also, we could determine which values are thresholds for change of behavior. While this is impossible, OrdEval algorithm uses the data sample and approximates this reasoning. For each respondent it selects its most similar respondents and makes inferences based on them. For example, to evaluate the effect an increase in a certain feature value would have on the overall satisfaction, the algorithm computes the probability for such an effect from the similar respondents with increased value of that feature. To get statistically valid and practically interesting results the overall process is repeated for a large enough number of respondents, and weighted with large enough number of similar respondents.

Feature (attribute) evaluation is an important component of many machine learning tasks, e.g. feature subset selection, constructive induction, decision and regression tree learning. Scores assigned to attributes during evaluation, also provide important information to the domain expert trying to get an insight into the problem domain. In this chapter we are interested in a subclass of feature evaluation, namely evaluation of conditionally strongly dependent ordinal attributes where each of the individual attribute's values may be dependent on other attributes in a different way. The problem of feature (attribute) evaluation has received a lot of attention in the literature. There are several measures for evaluation of attributes' quality. For classification problems the most popular are e.g. Gini index (Breiman et al., 1984), gain ratio (Quinlan, 1993), MDL (Kononenko, 1995), and ReliefF (Kononenko, 1994; Robnik-Šikonja & Kononenko, 2003). The first three are impurity based and measure quality of attribute according to the purity of class value distribution after the split on the values of that attribute. They evaluate each attribute separately, are not aware of the ordering of the attribute's values and cannot provide useful information for each individual value of the attribute. ReliefF on the other hand is context sensitive (by measuring how the attribute separates similar instances) and could be adapted to handle ordered attributes (by changing the definition of its similarity measure), but cannot provide information for each value separately and does not differentiate between the positive and negative changes of the attribute and their impact on the class value. By converting ordered nominal attributes into numeric ones we could use RReliefF, a regression version of ReliefF (Robnik-Sikonja & Kononenko, 2003) which, at the cost of assuming linear ordering, naturally handles positive and negative changes of an attribute, but cannot separate between positive and negative impact on a class value; also the extraction of information for each individual attribute's value is not possible in the RReliefF.

A typical approach in practical marketing research of customer (dis)satisfaction is to define a number of features of the product/service and then to conduct a survey on a representative sample of customers where the customers rate their satisfaction with each of the features and also express their overall (dis)satisfaction. While all types of captured answers (features) can be integrated with our approach, we are interested only in those questions of the survey which correspond to the product/service features. We consider them attributes in our data set and the overall (dis)satisfaction corresponds to the class value. The goal of feature analysis in marketing research is manifold:

1. Identify features which influence the overall (dis)satisfaction most,
2. Identify type of features: marketing research differentiates mostly between three types of important features:
    a. *Basic features* are taken for granted by customers. High score in these features does not significantly increase the overall satisfaction, while a low score usually causes dissatisfaction.
    b. *Performance features* are important features not taken for granted; they usually have a positive correlation with overall satisfaction: the higher the score the bigger the effect on the overall satisfaction,
    c. *Excitement features* usually describes properties of product/service which are normally not very important to the users, but can cause excitement (and boost in satisfaction) if the score is very high.
3. Identify those attribute values (thresholds) which have positive/negative impact on overall satisfaction, and
4. Identify typical behavior patterns of attribute values:
    a. *Upward reinforcement*: the value of a feature has a positive effect on overall satisfaction,
    b. *Downward reinforcement*: the value of a feature has a negative effect on overall satisfaction,
    c. *Anchoring*: the value of a feature acts as an anchor on overall (dis)satisfaction and prevents its change,

d. *Compensation* is a type of behavior characteristic for subsets of features, namely low score in one of the features is compensated with high score in one or more others.

There are many different feature evaluation algorithms in data mining which can evaluate, rank and/or select the most informative features (goal 1). For problems with highly dependent features as is predominantly the case in the marketing research, the most suitable heuristics are probably ReliefF (Robnik-Sikonja & Kononenko, 2003) and CM (Hong, 1997). Other goals (2-4) remain mostly untouched by current work in machine learning and data mining.

The motivation and contribution of this chapter is to demonstrate how OrdEval works, how its output can be visualized and adapted to include information relevant for practitioners. As a means for this we use a marketing context and present application of OrdEval on two costumer (dis)satisfaction studies.

## ORDEVAL APPROACH

OrdEval algorithm can be used for analysis of any data where the dependent variable has ordered values, meaning that it is also suitable for surveys where answers are given in the graded manner. The methodology uses conditional probabilities called 'reinforcement factors' as they approximate the upward and downward reinforcement effect the particular feature value has on the dependent attribute. For each value of the feature we obtain estimates of two conditional probabilities: the probability that the response value increases given the increase of the feature value (upward reinforcement), and the probability that the response value decreases given the decrease of the feature value (downward reinforcement). To take the context of other features into account, these probabilities are computed in the local context, from the most similar instances. The visualization of these factors gives clear clues about the role of each feature, the importance of each value and the threshold values. To understand the idea of OrdEval algorithm, the feature should not be treated as a whole. Rather we shall observe the effect a single value of the feature may have.

We use a notation where each of the $n$ learning instances $I$ is represented by an ordered pair $(\boldsymbol{x},y)$, where each vector of attributes $\boldsymbol{x}$ consists of individual attributes $A_i$, $i=1,...,a$, ($a$ is the number of attributes) and is labeled with one of the ordered class values $y_j$, $j=1,...,c$ ($c$ is the number of class values) $y_1<y_2<...<y_c$. Each discrete ordered attribute $A_i$ has values $1$ through $m_i$ ($m_i$ is the number of values of attribute $A_i$). We use $y(I_t)$ in a functional form when we refer to the class value of the $t$-th instance and $I_{t,i}$ when we refer to the value of the attribute $A_i$ for the $t$-th instance. We write $p(y_j)$ as the probability of the class value $y_j$.

To explain the idea of the approach we need some definitions. Let $R$ be a randomly selected observation and $S$ the observation most similar to it. Let $j$ be the value of the feature $A_i$ at observation $R$. We observe the necessary changes of response value and features ($A_i$ with value $j$ in particular) which would change $S$ to $R$. If these changes are positive (increase of response and/or feature values), let us define the following probabilities.

- $P(A_{i,j}^p)$ is a probability that $j$ (the value of feature $A_i$ at $R$) is larger than the value of feature $A_i$ at its most similar observation $S$. By estimating $P(A_{i,j}^p)$ we gather evidence of the probability that the similar observation $S$ has lower value of $A_i$ and the change of $S$ to $R$ is positive.
- $P(C_{i,j}^p A_{i,j}^p)$ is a probability that both the response as well as $j$ (the value of feature $A_i$ at $R$) are larger than the response and feature value of its most similar observation $S$. With $P(C_{i,j}^p A_{i,j}^p)$ we estimate the probability that positive change in both the response and $A_i$ value of similar instance $S$ is needed to get the values of $R$.

Similarly, for negative changes which would turn S into R (decrease of response and/or feature values), we define $P(A^n_{i,j})$ and $P(C^n_{i,j}A^n_{i,j})$. The outputs of the algorithm are two factors, upward and downward reinforcement, computed for each value of each feature. These factors measure the upward/downward trends exhibited in the data. The upward reinforcement of the *i*-th feature's value *j* is defined as

$$U_{i,j} = P(C^p_{i,j} \mid A^p_{i,j}) = \frac{P(C^p_{i,j}A^p_{i,j})}{P(A^p_{i,j})} \qquad (1)$$

This factor reports the probability that a positive response change is caused by the positive feature change. This intuitively corresponds to the effect the positive change in a feature's value has on the response. Similarly the downward reinforcement is defined as

$$D_{i,j} = P(C^n_{i,j} \mid A^n_{i,j}) = \frac{P(C^n_{i,j}A^n_{i,j})}{P(A^n_{i,j})}\ , \qquad (2)$$

and reports the effect the decrease of attribute's value has on the decrease of the class' value. The $U_{i,j}$ and $D_{i,j}$ factors are efficiently estimated by the OrdEval algorithm which we present below in a simplified form intended for easier comprehension.

*Input*: for each respondent a vector of feature values and the overall score
*Output*: $U_{i,j}$ and $D_{i,j}$ for all features *i* and their values *j*

1. *for* all features *i* and their values *j* initialize $A^p_{i,j}$, $A^n_{i,j}$, $C^p_{i,j}A^p_{i,j}$, $C^n_{i,j}A^n_{i,j}$ to 0
2. *for* pre-specified number of respondents
3. randomly select a respondent *R*
4. find k nearest respondents closest to *R*
5. *for* each closest respondent *S* and each features *i*
   update weights of $A^p_{i,j}$, $A^n_{i,j}$, $C^p_{i,j}A^p_{i,j}$, $C^n_{i,j}A^n_{i,j}$ as follows:
6. *if* feature value of *S* is lower than *j* increment $A^p_{i,j}$,
7. *if* both feature and overall score value of *S* are lower than *j* increment $C^p_{i,j}A^p_{i,j}$,
8. *if* feature value of *S* is higher than *j* increment $A^n_{i,j}$,
9. *if* both feature and overall score value of *S* are higher than *j* increment $C^n_{i,j}A^n_{i,j}$
10. *for* all features *i* and their values *j* compute
11. $U_{i,j} = C^p_{i,j}A^p_{i,j} / A^p_{i,j}$ and $D_{i,j} = C^n_{i,j}A^p_{i,j} / A^n_{i,j}$

The algorithm assumes that the cause of the differences in overall score are the differences in the attributes' values and gives these values some credit for that, but only if the sign of the differences in class and attribute is the same. It first sets counters of (co)occurring changes to zero (line 1). Than it randomly selects a respondent *R* (line 3) and searches for its k nearest respondents (line 4). For each of these most similar respondents it updates the counters for all the features depending on the overall scores and feature values of the randomly selected respondent and the near respondents (lines 5 - 9): if the feature value of the near instance is lower than the value of the random instance (line 6) then the change is positive and we update $A_{i,j}^{p}$, for the value *j* of the given feature *i* (*j* is the value of feature *i* for respondent *R*). If additionally the overall score of the similar respondent is lower than the score of the random respondent (line 7) then the change in both overall score and feature is positive and we update $C_{i,j}^{p}A_{i,j}^{p}$, for given feature *i* and its value of random respondent *j*. Similarly we do for negative changes in feature and overall score (lines 8-9). We repeat the whole process (lines 2 - 9) for a pre-specified number of iterations. Conservatively we can set this number to be equal to the number of respondents, but we get useful results even if we run only a few iterations (e.g., logarithm of the number of respondents). Finally the upward and downward enforcement factors for all the values of attributes are computed as conditional probabilities (lines 10-11).

The increments depend on the values of attribute $A_i$ at the random instances *R* and its near instance *S*. The simplest form of update function *w(R,S)* (see Eq. (3)) takes into account only the number of nearest instances *k*. The idea is to average the results of k nearest instances.

$$w(R,S) = \frac{1}{k} \qquad (3)$$

In our work we use value *k=10* as this is the default value of most k-nearest neighbor classification studies. Depending on the nature of the problem other values could be more suitable, but this is not the topic of this work. For discussion see (Duda et al., 2001). A more sophisticated version of the updated function takes also the distance between the instances into account: closer instances should have greater impact. Exponentially decreasing weighted contribution of instances ranked by distance is recommended by Robnik-Šikonja and Kononenko (2003). See this reference for detailed explanation of this issue.

Missing entries which frequently occur in these types of problems can simply be excluded from computation. The similarity of instances is computed as Manhattan distance over all the attributes:

$$d(I_t, I_u) = \sum_{i=1}^{a} \mathrm{diff}(A_i, I_t, I_u). \qquad (4)$$

The single attribute distance for two instances $I_t$ and $I_u$ is computed with the function diff($A_i$,$I_t$,$I_u$). For ordinal attributes, when we do not have better domain knowledge available we can assume linear ordering:

$$\mathrm{diff}(A_i, I_t, I_u) = \frac{|I_{t,i} - I_{u,i}|}{m_i - 1}, \qquad (5)$$

where $m_i$ is the number of values of feature $A_i$.

The computational complexity of the algorithm is of the order $O(m \cdot n \cdot a)$. The main computational burden within each of the *m* iterations is the search for the nearest instances (line 4), for which we have to compute the distances to all the instances in $O(n \cdot a)$ steps. If the number of features is low we can reduce this computation by the use of smart data structures (for example k-d trees or R-trees) or we can investigate performance of approximate k-nearest neighbors algorithms (Duda et al., 2001). Each iteration of the algorithm is independent and here we see a path to parallelization of the algorithm.

Our algorithm does not cover all the information which could be extracted with the help of class and feature value changes in the nearest neighbor context. Important aspect could also be hidden in the amount of positive and negative changes resulting from no changes in attribute's value. This could be a useful hint about the coherency in the data, noise, overall importance of the attribute, as well as the amount of dependency between the attributes. Also important is the anchoring effect uncovered by our algorithm, but this remains as further work.

## Illustrative Example

To show the behavior and usability of the algorithm we first define a simple artificial problem which is motivated by the Behavioral Decision Theory, stating that there are several distinct manners according to which marketing stimuli can be used during the formation of product attitude (Einhorn & Hogarth, 1981).

Our data set is described by six important and two irrelevant features. The important features correspond to different feature types from the marketing theory: two basic features ($B_{weak}$ and $B_{strong}$), two performance features ($P_{weak}$ and $P_{strong}$), two excitement features ($E_{weak}$ and $E_{strong}$), and two irrelevant features ($I_{uniform}$ and $I_{normal}$). The values of all features are randomly generated integer values from 1 to 5, indicating for example score assigned to each of the features by the survey's respondent. The dependent variable for each instance (class) is the sum of its features' effects, which we scale to the uniform distribution of integers 1-5, indicating, for example, an overall score assigned by the respondent.

$$C = b_w(B_{weak}) + b_s(B_{strong}) + p_w(P_{weak}) + p_s(P_{strong}) + e_w(E_{weak}) + e_s(E_{strong})$$

The effects of attributes are as follows.

- Basic features are taken for granted by customers; a high score in these features does not significantly increase the overall score, while a low score has a decreasing effect on dependent variable. We define two variants of basic features, one with weaker and another with stronger negative impact:

$$b_w(A) = \begin{cases} -2 & ;A \le 2 \\ 0 & ;A \ge 3 \end{cases}, \qquad b_s(A) = \begin{cases} -4 & ;A \le 3 \\ -2 & ;A = 4 \\ 0 & ;A = 5 \end{cases}.$$

- Performance features have a positive correlation with the overall score: the higher the value of the attribute the bigger the effect on the overall score. We define the performance effects as

$$p_w(A) = \begin{cases} -3 & ;A = 1 \\ -2 & ;A = 2 \\ 0 & ;A = 3 \\ 2 & ;A = 4 \\ 3 & ;A = 5 \end{cases}, \qquad p_s(A) = \begin{cases} -5 & ;A = 1 \\ -3 & ;A = 2 \\ 0 & ;A = 3 \\ 3 & ;A = 4 \\ 5 & ;A = 5 \end{cases}.$$

- Excitement features describe properties of product/service which are normally not very important to the users, but can cause excitement if the score is very high. We define two grades of excitement effect as

$$e_w(A) = \begin{Bmatrix} 0 & ; A \le 4 \\ 1 & ; A = 5 \end{Bmatrix}, \qquad e_s(A) = \begin{Bmatrix} 0 & ; A \le 4 \\ 4 & ; A = 5 \end{Bmatrix}.$$

We generated 1000 instances for this data set. While the value distribution and the independence of features are unrealistic, note that we have experimented also with more realistic distributions as well as with different types of correlation, but the results and conclusions remain unchanged. The upward and downward reinforcement factors the OrdEval algorithm returned for this data set are probabilities whose direct interpretation and analysis is of course possible, but visualization makes it easier.

The slope visualization proposed by Robnik-Sikonja and Vanhoof (2007) (upward and downward reinforcement are represented with the steepness of the line segment between two consecutive feature values) is unusual for marketing research practitioners and, as we argue below, does not convey all the information necessary for this specific field. We therefore propose a marketing friendly visualization of the OrdEval results on Figure 1, which contains results for each feature separately.

*Figure 1. Visualization of reinforcement factors and their confidence intervals on the problem with different types of features*

The eight subgraphs are a sort of bar charts with addition of confidence intervals. For each graph a left-hand side contains downwards reinforcements for each feature score separately. Upwards reinforcement factors for all the scores are represented on the right-hand side of each graph. Before we explain the results let us give a motivation for box-and-whiskers graphs on top of each reinforcement bar.

There are two problems with these reinforcement factors in general and also when used in marketing:

- Imbalanced value distribution: it is quite common that for certain features some scores are almost non-existent (e.g., extremely low score of a basic feature is very rare - such a customer, would probably change the supplier), and also the reverse might be true, namely on a scale 1-5 it is not uncommon that almost all the scores are 4 and 5. Such imbalance also has consequences for reinforcement factors, since the probability of the increased/decreased overall score might be an artifact of the skewed distribution of values.
- Lack of information about significance of the reinforcement factors: the user does not know what expected range of a certain reinforcement factor is and weather the computed score is significantly different from the uninformative feature.

To solve both problems we compute confidence intervals for each reinforcement factor. Since we cannot assume any parametric distribution and have to take the context of a similar respondent into account we construct bootstrap estimates and form confidence intervals based on them (Efron & Tibshirani, 1993). We proceed as follows:

1. For each feature we construct e.g. n=200 features with bootstrap sampled values from the original feature (alternatively the values can be randomly shuffled), we call these features normalizing features,
2. When searching similar respondents we only take original features into account, but we estimate also the reinforcement factors of randomly constructed features,

3. For each reinforcement factor $U_{i,j}$ and $D_{i,j}$ (upward and downward reinforcement for each value of each feature) we perform a statistical testing based on bootstrap estimates.
    a. The null hypothesis states that the reinforcement factor is uninformative, i.e., it is equal to the median of its random normalizing features
    b. The alternative hypothesis is one-sided, as we are interested if the reinforcement of the original feature is larger than the random normalizing reinforcement
    c. Set fixed confidence level, e.g. $\alpha$=0.05
    d. Sort the reinforcement factors of random normalizing features in ascending order
    e. If the reinforcement factor of the original feature is larger than $n(1- \alpha)$th sorted factor we can reject the null hypothesis, and assume that the computed reinforcement contains significant information
4. The sorted reinforcement factors are the source of information for box-and-whiskers plot: the box is constructed from the 1st and 3rd quartile, middle line is median, while the whiskers are $100\alpha/2$ and $100(1- \alpha )/2$ percentiles (e.g. 2.5 and 97.5 percentiles) giving the borders of confidence interval (e.g., 95% confidence interval).

In Figure 1, reinforcement factors reaching beyond the box-and-whiskers therefore contain significant information. Since the way we construct confidence intervals is not sensitive to the number of instances, these intervals are valid even for low number of scores. We can observe that the algorithm has captured the important landmarks of the features:

- For performance features $P_{weak}$ and $P_{strong}$ (two graphs in the top row) all the upward and downward reinforcements are significant, and the relative length of the bars is roughly proportional to the difference between impacts of the values,
- For basic feature $B_{weak}$ (left-hand graph in the second row) the thresholds at values 2 and 3 (increasing feature from 2 to 3 strongly increases the overall score, and decreasing this feature from 3 to 2 strongly decreases the overall score),
- For basic attribute $B_{strong}$ (right-hand graph in the second row) the upward thresholds at values 3 and 4 and downward reinforcement thresholds 4 and 5,
- For excitement features $E_{weak}$ and $E_{strong}$ (third row graphs) the jump from 4 to 5 and back is detected, in upward and downward enforcement, respectively. The reinforcements are larger for $E_{strong}$ as expected,
- Irrelevant random features $I_{uniform}$ and $I_{normal}$ have no significant values (bottom row).

Note that only the reinforcement for the thresholds we have defined, are significantly larger than the boundaries of confidence intervals defined by the normalization features.

The properties of the used approach relevant to our study in particular, and in more general terms, to the analysis of arbitrary survey at large, are manifold. Firstly, there is substantial *context sensitivity*. Typically the features are highly conditionally dependent upon the response and have to be evaluated in the context of other features. OrdEval is intrinsically contextual and assumes neither independence nor some fixed distribution of the features. The context of other features is handled through the distance. By using different distance measures and different features in the calculation of the distance, we are even in a position to use different contexts, e.g., we could use some background socio-economic information to calculate the similarity of respondents. Secondly, there is the *ability to handle ordered features and ordered response* and to use the information the ordering contains. The order of attribute's values contains information which is comparable but not the same as values of numerical features, e.g., values poor, good, very good, excellent are ordered in expressing certain attitude but this ordering is not necessarily linear. Thirdly, we have *awareness of the meaning implied by the ordering* of the answers and the positive (negative) correlation of changes between feature values and the response (e.g., if the value of the feature increases from poor to good, we have to be able to detect both positive and negative correlation to the change of the overall response value). Fourthly, OrdEval has the *ability to handle each value of the feature separately*, e.g., for some features the value of good and very good have identical neutral impact on the response, value poor may have a strong negative, and value excellent a highly positive impact. We are able to observe and quantify each feature's values separately and thereby identify important thresholds. Next to that, *visualization* of the output allows experts to use it as a powerful exploratory data analysis tool, e.g., to identify type of features and the impact of their individual values. Also, *the output is in the form of probabilities*. Probability theory is commonly used and therefore the results in form of probabilities are comprehensible and interpretable by a large audience and can also be used operationally. Finally, we have *fast computation* and *robustness to noise and missing values*. A study of the family of the algorithms similar to OrdEval has shown that feature evaluation is possible and reliable even for extremely noisy data (Robnik-Sikonja & Kononenko, 2003). Table 1 summarises the benefits and disadvantages of utilising various methods of data collection.

*Table 1. Data collection techniques*

| Approach | Benefits | Disadvantages |
|---|---|---|
| Paper based survey | Hard copy so highly visible.<br>Traditional and low barrier to use.<br>More user friendly.<br>No PC access required<br>No PC skills required | Difficulty of re-entering names generated from name generator question<br>Expensive due to mailing costs.<br>Labour intensive reminder system.<br>Social network questionnaires in this format are long & cumbersome |
| Online Survey | Cost efficient way to distribute survey.<br>Can re-use data from earlier questions (name generator) in later questions to avoid need for re-entry.<br>Software can facilitate question skipping based on answers given to reduce participant burden<br>Ease of introducing reminder system.<br>Respondent can fill in survey over number of days by constant saving.<br>Data entered in electronic form which facilitates analysis stage. | Possible barrier of the technology.<br>Not as visible as paper based survey. |
| Interviews | Facilitates correct understanding of questions.<br>Allows collection of 'deeper' information.<br>Ensures all questions are answered. | Can be more repetitive & more time consuming than respondent self-completion<br>Expensive due to travel costs.<br>Time pressure to respond in the face-to-face situation can reduce respondents' time to think & result in forgetting network members |

## Case Study: Customer Satisfaction Analysis in Marketing

Over the last forty years, consumer (dis)satisfaction has taken a prominent position in the marketing research literature (e.g. Anderson, 1973; Anderson & Sullivan, 1993; Cardozo, 1965; Churchill & Surprenant, 1982). This attention is justified since consumer (dis)satisfaction (directly or indirectly) impacts upon repurchase intention (Szymanski & Henard, 2001), consumer retention (Anderson, 1994; Mittal and Kamakura, 2001) and eventually upon firm performance (Anderson et al., 1994). Consumer (dis)satisfaction is a summarizing response that results from a consumer's post-consumption cognitive and affective evaluation of a product or service performance given pre-purchase expectations (Anderson and Sullivan, 1993; Oliver, 1993; Tse & Wilton, 1988).

Focusing on the antecedents of consumer (dis)satisfaction, two main issues dominate today's discourse: the expectancy-disconfirmation theory and the nature of the relationship between consumer (dis)satisfaction and its antecedents.

First, the expectancy-disconfirmation paradigm is a dominant framework for explaining consumer (dis)satisfaction (Oliver, 1997; Szymanski & Henard, 2001). In its basic format, the model proposes that consumers' overall (dis)satisfaction response is the result of two cognitive processes (Oliver, 1997). In the first, consumers form pre-purchase expectations on the performance of a product or service. In the second, consumers evaluate the actual performance of the product and compare this perceived performance to their expectations. If performance meets expectations, consumers experience confirmation of their expectations. If performance is greater than expected, they experience positive disconfirmation; if performance is less than expected, consumers experience negative disconfirmation (see Oliver (1997) and Yi (1991) for reviews).

Secondly and related to the first is the debate on the nature of the relationship between consumer (dis)satisfaction and its antecedents. Initially, the effects of the antecedents and in particular of attribute-level performance on consumer (dis)satisfaction were assumed linear and symmetric (Mittal et al., 1998; Sethi & King, 1999; Spreng et al., 1996). Only recently, marketing scholars have questioned this double assumption on the basis of economic and psychological theory as well as on a better empirical insight in the satisfaction response function (Anderson & Sullivan, 1993).

The presented attribute evaluation method attempts to extend the knowledge on the relationship between consumer (dis)satisfaction and its main antecedents. More specifically, we try to quantify and visualize the relationship between attribute-level (dis)satisfaction and overall (dis)satisfaction.

We report performance of our methods on one recent business-to-business (B2B) and one recent consumer-to-business (C2B) customer satisfaction study. For the business-to-business study the product involved is a high-tech product. Requirements are specified by the customer; the product is produced and delivered on demand. The whole process from order to delivery can take two or three months. The database provides the satisfaction scores of customers, who are active and have on-going orders. They reported their (dis)satisfaction with 11 product/service attributes as well as their overall satisfaction with the product. Overall satisfaction and attributes were measured on a 5-point scale. This data set is small (less than 100 records).

The consumer-to-business study is based on a study of a main European player in the entertainment sector. The survey is hierarchically organized. Satisfaction has been measured as general (overall satisfaction), on different dimensions (like personnel, administration, communication, etc.) and on aspects of the dimensions (like personnel friendliness, clearness of invoice, etc.). Dimensions and aspects are measured on a 10-point scale. The data set contains over 4000 instances.

Due to confidentiality we cannot go into details or give all the results. Therefore we will give some examples from both data sets. The chosen attributes/dimensions are attributes/dimensions that occur in most customer satisfaction data sets.

We report on the findings and types of behavior our algorithm can discover and give some relations to marketing literature. The visualization proposed by Robnik-Šikonja and Vanhoof (2007) conveys the information with the slope of the line segments between attribute values. The upward and downward reinforcement numbers represent the steepness of the line segment between two consecutive feature values (coefficient of the straight line between the two values) e.g., 0 is horizontal line, 1 corresponds to 45 degrees angle (the maximum), 0.5 ≈ 26.6 degrees angle, etc. We visualize $U_{ij}$ and $D_{i,j}$ for all the values of one attribute in a single graph on the left-hand side of Figure 2. Feature values are displayed on horizontal axis, and the cumulative enforcement (sum of reinforcements up to the particular value) on the vertical axis. Upward reinforcement factors cover the upper part of the graph and downward reinforcement factors reside in the lower part of the graph. Arrows indicate the direction of the change. Beside slopes note also the total cumulative height of the slopes as it is indicative of the overall importance of the feature.

For a real world data set it is of course irrational to expect so clear distinctions between features as evident for our toy data in Figure 1. In reality we find the same characteristics as predicted by theory (basic, performance, and excitement features) but less clearly expressed, or expressed only for some, usually most frequent, values.

We first look at the results obtained for "price" (Figure 2) from C2B study. The reinforcement factors are presented on the left-hand side with the slope of line segment, and on the right hand side with bar charts. If we only look at the left-hand side we see a clear picture of performance feature and the managerial conclusion would be that for all costumers the better the satisfaction with price the better the overall satisfaction, so with price we can clearly regulate the overall satisfaction. The right-hand side graph gives much more precise information. First, the human eye is not sensitive enough to the angle of the line segment slope; therefore it is difficult to see even quite large differences. Additional benefit of confidence intervals is also clearly visible, namely for upward reinforcement only increases from values 1, 2, 3, and 4 are significantly larger (at 95% level) than the score obtained by random permutation of values. For downward reinforcement the significant values are 7, 6, and 5. The managerial consequences are clear: if management wants to increase the satisfaction of the customers the price reduction will have effect only on the least satisfied customers (scores 1, 2, 3, 4). On the other hand price increase will decrease overall satisfaction for a group of respondents who claim medium price satisfaction. Similar conclusions are possible for other attributes and allow much more focused and precise managerial decisions as with classical statistical approaches, e.g., linear regression.

*Figure 2. The results for "price" in C2B data set. The left-hand side visualizes the reinforcement factors with the slope of line segment, and the right hand side gives additional information with bar charts and confidence intervals.*

Besides “price” several other attributes in B2C and B2C study can also be classified as performance attributes, e.g., "information about promotions", or “communication", where for several values an increase (decrease) of the attribute level influences satisfaction (dissatisfaction).

Several attributes show a very similar pattern to a basic attribute. The example given is "product quality" (Figure 3) from B2B study.

*Figure 3. The results for "product quality" in B2B data set.The left-hand side shows reinforcement factors with confidence intervals, while the right-hand side shows the same information, but significant values are indicated with a different shade.*

A basic attribute behaves like a threshold or stepwise function: it creates (dis)satisfaction when (not) fulfilled. For example ‘quality of the product’ (Figure 3) is a basic requirement that obtains the quality level 3 or higher. Only these values are presented in the study which is not surprising (customers dissatisfied with the product quality at this level would change the supplier). We can observe that only increase from 3 to 4 significantly influences satisfaction and a decrease from 5 to 4 significantly influences dissatisfaction.

Prospect theory in general and its assumptions of loss aversion and diminishing sensitivity in particular (Einhorn & Hogarth, 1981; Mittal & Kamakura, 2001) proposes an asymmetric S-shaped relationship between attribute-level performance and overall (dis)satisfaction. Indeed, it has been observed that the marginal contribution of attribute-level performance on overall (dis)satisfaction decreases with its size and that losses have more impact than gains. This is confirmed by evidence and theory on memory accessibility: negative information is more perceptually salient, is given more weight, and creates a stronger response than positive information (Mittal & Kamakura, 2001; Peeters & Czapinski, 1990). Our results show higher values for the downward reinforcement than for upward reinforcement. All these figures demonstrate clearly the asymmetric and non-linear nature of the relationship between consumer (dis)satisfaction and the attribute under consideration. Marketing managers are specifically interested in the height and the shape of the curves and the position of the breakpoints.

## FUTURE RESEARCH DIRECTIONS

It is our belief, that the ordered attribute evaluation can be used in fields other than marketing. The algorithm and its visualizations can be useful in any survey analysis where the answers are graded. So far we have used only part of information hidden in the difference between class and feature values of similar instances. Other effects in marketing (such as anchoring) and new applications in other fields may demand definition of additional factors and development of novel visualization techniques.

For example, survey data collection is one of the most important sources of socio-demographic data and its quality is of key importance for informed policy-decision making. The design of survey questionnaires (visual appearance, structural organization, and wording) for collecting socio-demographic data is one of the most important factors affecting the quality of the data. We believe that OrdEval could be efficient in testing and evaluating the wording of survey questions based on experimental or post-survey data in the socio-demographic research. Due to OrdEval properties it is possible to adapt this analytical approach to test and evaluate survey questionnaires where many of the key survey variables are using ordinal scales.

## CONCLUSION

OrdEval algorithm exploits the information hidden in the ordering of class and attribute values and their inherent correlation. Based on nearest neighbor paradigm and probability theory the algorithm is context sensitive, able to handle ordered attributes and ordered classes, aware of the information the ordering contains, able to handle each value of the attribute separately, and provides output which can be effectively visualized. The visualizations we developed turned out highly useful in our marketing research case study. From a data mining point of view the paper has adapted a general methodology for analysis of ordered data to the specifics of marketing. OrdEval algorithm possesses also other favorable properties like output in the form of probabilities, fast computation, robustness to noise and missing values, and a possibility of parallelization. Additionally it is possible to efficiently compute confidence intervals for reinforcement factors. For example imbalanced value distribution is quite a common phenomenon but it has severe consequences for reinforcement factors, since the probability of the increased/decreased overall score might be an artifact of the skewed distribution of values. Another such obstacle is information about the significance of the reinforcement factors: the user does not know what the expected range of a certain reinforcement factor is and weather computed score is significantly different from the uninformative feature. By computing distribution independent confidence intervals we provide information on reliability of the reinforcement scores which give them practical importance and enables confident decision making. Additionally the proposed visualization of the reinforcement factors enables detection of (non)linearity, (a)symmetry, threshold values and significance of the results.

This chapter also has technical implications for academic research on costumer (dis)satisfaction. Extracting the kind of knowledge we discussed in the present study is not self-evident. In marketing research, the potential of the ordered attribute evaluation to unravel the decision-making heuristics of customers when ’deciding’ on a certain level of (dis)satisfaction seems to outperform that of more traditional statistical models. This is due to the power of the method to allow for non-linear and asymmetric effects as well as to the fact that researchers should not a priori postulate the roles the different attributes will take. Although the algorithm appears analytically complex, it may yield parsimonious results. This paper illustrates and confirms earlier advice that managers should identify the ‘optimal’ performance level for each attribute. The goal should be to optimize, not to maximize attribute-level performance at a level where the payoff in terms of overall customer (dis)satisfaction is maximized. This optimal level can be determined by analyzing the different figures offered by our method. As such, the OrdEval algorithm can be a valuable technical contribution to the analysis of this particular task.

## REFERENCES


Anderson, E. W. (1994). Cross-category variation in customer satisfaction and retention. *Marketing Letters*, *5*, 19–30.

Anderson, E. W. and Sullivan, M. (1993). The antecedents and consequences of customer satisfaction for firms. *Science*, *12*(2), 125–143.

Anderson, E. W., Fornell, C. & Lehmann, D. (1994). Customer satisfaction, market share and profitability: Findings from Sweden. *Journal of Marketing*, *58*, 63–66.

Anderson, R. E. (1973). Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *Journal of Marketing Research*, *10* (February), 38–44.

Breiman, L. , Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984) . *Classification and Regression trees*. Belmont, CA: Wadsworth, Inc.

Cardozo, R. N. (1965). An experimental study of customer effort, expectations, and satisfaction. *Journal of Marketing Research*, *2*(August), 244–249.

Churchill, G. A. J. & Surprenant, C. (1982). An investigation into the determinants of consumer satisfaction. *Journal of Marketing Research*, *19*, 491–504.

Duda, R. O., Hart, P. E. & Stork, D. G. (2001). *Pattern Classification (2nd Ed.)*. New York: John Wiley & Sons.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to Bootstrap*. New York: Chapman & Hall.

Einhorn, H. J. & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgement and choice. *Annual Review of Psychology*, *32*, 53–88.

Hong, S. J. (1997). Use of contextual information for feature ranking and discretization. *IEEE transactions on knowledge and data engineering*, *9*(5), 718–730.

Kononenko, I. (1994). Estimating attributes: analysis and extensions of Relief. In L. De Raedt & F. Bergadano, (Eds.), *Machine Learning: ECML-94,* (pp. 171–182). Berlin: Springer Verlag.

Kononenko, I. (1995). On biases in estimating multi-valued attributes. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'95)*, (pp. 1034–1040). San Francisco: Morgan Kaufmann,.

Mittal, V. & Kamakura, W. (2001). Satisfaction, repurchase intention, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, *38*(February), 131–142.

Mittal, V., Ross, W. J. & Baldasare, P. (1998). The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing*, *62*(January), 33–47.

Oliver, R. L. (1993). Cognitive, affective and attribute bases of the satisfaction response. *Journal of Consumer Research*, *20*, 418–430.

Oliver, R. L. (1997). *Satisfaction: A Behavioral Perspective on the Consumer*. New York: McGraw-Hill.

Peeters, G. & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effect. *European Review of Social Psychology*, *1*, 33–60.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kaufmann.

Robnik-Šikonja, M. & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning Journal*, *53*, 23–69.

Robnik-Šikonja, M. & Vanhoof, K. (2007). Evaluation of ordinal attributes at value level. *Data Mining and Knowledge Discovery*, *14*, 225–243.

Sethi, V. & King, R. (1999). Nonlinear and noncompensatory models in user information satisfaction measurement. *Information Systems Research*, *10*(1), 87–96.

Spreng, R. A., McKensey, S., & Olshavsky, R. (1996). A reexamination of the determinants of consumer satisfaction. *Journal of Marketing*, *60*(July), 15–32.

Szymanski, D. M. & Henard, D. (2001). Consumer satisfaction: A meta-analysis of the empirical evidence. *Journal of the Academy of Marketing Science*, *29*(1), 16–35.

Tse, D. K. & Wilton, P. (1988). Models of consumer satisfaction formation: An extension. *Journal of Marketing Research*, *25*(May), 204–212.

Wikipedia (n.d.) Shareholder. *Wikipedia.* Retrieved from: https://en.wikipedia.org/wiki/Shareholder

Wikipedia – Stockholders (n.d.) Stockholders. *Wikipedia.* Retrieved from: https://en.wikipedia.org/?title=Stockholders&redirect=no

Yi, Y. (1991). A critical review of consumer satisfaction. In V. A. Zeithaml, (Ed.), *Review of Marketing*, (pp. 68–123). Chicago: American Marketing Association.

## ADDITIONAL READING

Ben-David, A. (1995). Monotonicity maintenance in information-theoretic machine learning algorithms. *Machine Learning*, *19*, 29-43.

Breiman, L. (2001). Random Forests. *Machine Learning Journal*, *45*, 5-32.

Cardoso, J. S., & da Costa, J. E. P. (2007). Learning to classify ordinal data: The data replication method. *Journal of Machine Learning Research*, *8*, 1393-1429.

Demšar, J. (2006). Statistical comparison of classifiers over multiple datasets. *Journal of Machine Learning Research*, *7*, 1-30.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*, 1157-1182.

### KEY TERMS AND DEFINITIONS

**Ordered Attribute**: An attribute with nominal, but ordered values, for example, increasing levels of satisfaction: low, medium, and high.

**Evaluation of Ordered Attributes**: For ordered attributes the evaluation procedure should take into account their double nature: they are nominal, but also behave as numeric attributes. So each value may have its distinct behavior, but values are also ordered and may have increasing impact.

**Attribute Evaluation:** A data mining procedure which estimates the utility of attributes for given task (usually prediction). Attribute evaluation is used in many data mining tasks, for example in feature subset selection, feature weighting, feature ranking, feature construction, decision and regression tree building, data discretization, visualization, and comprehension.

**Context of Attributes**: In a given problem the related attributes, which interact in the description of the problem. Only together these attributes contain sufficient information for classification of instances. The relevant context may not be the same for all the instances in given problem.

**Non-Myopic Attribute Evaluation**: An attribute evaluation procedure which does not assume conditional independence of attributes but takes context into account. This allows proper evaluation of attributes which take part in strong interactions.

**Feature Subset Selection**: Procedure for reduction of data dimensionality with a goal to select the most relevant set of features for a given task trying not to sacrifice the performance.

**Feature Weighting**: Under the assumption that not all attributes (dimensions) are equally important feature weighting assigns different weights to them and thereby transforms the problem space. This is used in data mining tasks where the distance between instances is explicitly taken into account.