Information Sciences 198 (2012) 118–129

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

A rule-based method for identifying the factor structure in customer satisfaction

Amir Ahmad [a,*], Lipika Dey [b], Sami M. Halawani [a]

[a] *Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh, Saudi Arabia*
[b] *Innovation Labs, Tata Consultancy Services, New Delhi, India*

ARTICLE INFO

Article history:
Received 17 March 2008
Received in revised form 14 July 2011
Accepted 25 February 2012
Available online 7 March 2012

Keywords:
Market research
Customer satisfaction
Three-factor theory
Feature value importance
Categorical features

ABSTRACT

The analysis of customer satisfaction datasets has shown that product-related features fall into three categories (i.e., basic, performance, and excitement), which affect overall satisfaction differently. Because the relationship between product features and customer satisfaction is characterized by non-linearity and asymmetry, feature values are studied to understand the characteristics of a feature. However, existing methods are computationally expensive and work for ordinal features only. We propose a rule-based method that can be used to analyze data features regarding various characteristics of customer satisfaction. The inputs for these rules are derived by using a probabilistic feature-selection technique. In this feature selection method, mutual associations between feature values and class decisions in a pre-classified database are computed to measure the significance of feature values. The proposed method can be used for both types of features: ordinal and categorical. The proposed method is more computationally efficient than previously recommended methods. We performed experiments on a synthetic dataset with known characteristics, and our method correctly predicted the characteristics of the dataset. We also performed experiments with a real-housing dataset. The knowledge extracted from the dataset by using this method is in agreement with the domain knowledge.

1. Introduction

Market research is the process of designing, gathering, analyzing, and reporting information that may be used to solve a specific marketing problem [11]. Market research can help companies understand their customers so that they can modify their strategy to attract more customers.

Estimating the relationship between product features and customer satisfaction and customer dissatisfaction is fundamental to market research. The use of symmetric linear-functional forms is popular in market research [10,17,49]. Symmetric functions assume that both positive and negative performance have equal impacts on customer satisfaction; however, the current literature [4,31,32] suggests that the relationship between performance and customer satisfaction is in fact characterized by non-linearity and asymmetry.

The three-factor theory of customer satisfaction [24] is gaining popularity in market research problems [4,22,31,44]. The three-factor structure of customer satisfaction can be described as a combination of the following three factors:

(a) *Basic factors* – Customers take these basic factors for granted. Although these factors contribute very little towards customer satisfaction, the absence of these factors ultimately leads to dissatisfaction.

* Corresponding author. Tel.: +966 551346386.
E-mail address: amirahmad01@gmail.com (A. Ahmad).

0020-0255/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2012.02.056

(b) *Performance factors* – These factors have both positive and negative effects on customer satisfaction. If they are provided, these factors positively contribute towards customer satisfaction, and if they are not provided, then these factors negatively affect customer satisfaction.
(c) *Excitement factors* – These factors do not contribute towards customer dissatisfaction if they are not provided, but if they are provided, they positively contribute towards customer satisfaction.

Identifying these features is important so that business managers can set their priorities to maximize customer satisfaction. The three-factor theory [24] has been applied to different fields, such as web-based learning [27] and price formation [35].

As the amount of data increases, the demand for efficient algorithms to analyse the data also increases. Data mining is "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" [13]. Data mining techniques have been successfully used to address different market research problems [6]. Identifying target groups for marketing purposes is one of the most common applications of data mining. Association rules [2,3,43], clustering algorithms [14], and finding the k-nearest neighbors [34] are some of the data mining techniques that have been successfully applied to this problem. Relationship marketing (RM) (also known as customer relationship management (CRM)) is a "strategy to attract, retain and enhance customer relationships" [5]. The analysis of customer behavior patterns is an important part of CRM. Berry and Linoff [6] describe data mining techniques for detecting customer behavior patterns. Several data mining techniques, such as clustering [41,48], analysis of feature values [42], text mining [46], agent-based models [21], dominance-based rough set approach [29], and concept-based learning [15,16], have been used to analyse customer data. However, not many methods are available for identifying the three-factor structure in customer satisfaction. The methods proposed by Mittal and Kamakura [32], and Sikonja and Vanhoof [42] are very popular for this purpose.

Sikonja and Vanhoof [42] proposed that the three-factor structure problem can be studied by using a feature selection method, RELIEF [25]. RELIEF [42] computes the importance of feature values by identifying the feature values that have either a positive or negative effect on overall satisfaction. The terms "upward reinforcement" and "downward reinforcement" were also introduced in the paper [42]. Upward reinforcement defines the increase in probability of the satisfaction value as the feature values increase. In the same way, downward reinforcement defines the decrease in probability of the satisfaction value as the feature values decrease. The computation was based on finding contextual information by using the k-nearest neighbors in each member of a set S of training points. For each point in the set S, n distance calculations (where n is the number of data points in the dataset) are needed, which makes this process rather slow.

The method proposed by Sikonja and Vanhoof [42] paved the way to analyze customer satisfaction and dissatisfaction from a different perspective. However, the proposed method worked for ordinal features only and could not handle categorical features, such as color and gender. A qualitative variable Y is considered categorical if its range bears no internal structure [9]. Thus, for the two categorical feature values x and y, one can only distinguish between two alternatives, $x = y$ and $x \neq y$. Market datasets are often expressed using such categorical features. Paulssen and Sommerfeld [36] proposed the use of dummy numeric codes to convert from categorical to numeric feature values. However, this process often leads to a loss in critical information.

The application of a feature selection method for the three-factor theory [42] motivated us to develop a method using a feature selection technique [1] for identifying the three-factor structure in customer satisfaction data that can overcome the insufficiencies of the method proposed by Sikonja and Vanhoof [42]. As most of the business managers are not artificial intelligence (AI) experts, it is important to develop a system that is easily understandable to non-AI experts. Rule-based methods are popular because they are easily interpretable. We propose a rule-based method for the identification of the three-factor structure of customer satisfaction. The inputs of these rules are generated by using the feature selection method proposed by Ahmad and Dey [1]. We summarize our contributions in the following points:

(a) Because the proposed method is a rule-based method, the results are easily interpretable by domain experts.
(b) The proposed method is computationally more efficient than the other popular method [42].
(c) The proposed method is more general compared to the other methods [4,31,42] because they can be applied to ordinal datasets only, whereas the proposed method can be applied to categorical datasets.
(d) The proposed method does not assume any underlying statistical distribution, as do regression methods; hence, it avoids model-misspecification.

The rest of the paper is organized as follows. In Section 2, we describe some related works. In Section 3, we discuss the probabilistic method proposed by Ahmad and Dey [1] to compute the significance of feature values and to present the proposed rule-based method. In Section 4, we show the effectiveness of the proposed method by applying it to a synthetic customer satisfaction dataset, which is suggested in [42], and to a housing dataset. Section 5 presents our conclusions and discusses possible future extensions of our work.

2. Related work

Feature selection is an important aspect of data mining that can substantially improve the results. There are different feature selection techniques that can be used to rate or choose the important features from given data. Blum and Langley [8] categorized the feature selection techniques for classification into three basic approaches. In the first approach, the *embedded approach*, a basic induction method is used to add or remove features from the concept description space in response to prediction errors on new instances [40]. The second approach, the *filtering approach*, explores various subsets of features to find an optimal subset that preserves the classification knowledge [33]. The third approach, the *wrapper technique* [23], evaluates alternative feature sets by running an induction algorithm on the training data and by using the estimated accuracy of the resulting classifier as a metric. Kira and Rendell [25] proposed a different approach for feature selection, the RELIEF algorithm, which weights each feature based on the ability of the feature to distinguish among the classes. The features whose weights exceed a user-defined threshold are retained as relevant. Kononenko [26] suggested the use of k-nearest neighbors to increase the reliability of the probability approximation. They also suggested that RELIEF can be extended to work with multiple sets more efficiently.

Rough set theory [37–39] is a mathematical tool to deal with uncertainty and vagueness of decision systems and this has been successfully applied to the feature selection problem [7,28,45]. In a rough set framework, a *reduct* is defined as the minimal subset of attributes having the same discernibility power as the original set of attributes. Pawlak and Skowron [38] proposed the use of *core*, which is the intersection of all reducts as a collection of important features.

Swiniarski and Skowron [45] combined rough set technologies and dimensionality reduction techniques, such as principal components analysis (PCA), to find the relevant features. Hu et al. [19] developed a simple but efficient feature selection technique that can work without discretizing the data, which is traditionally required by rough set-based methods. Hu et al. [20] then proposed a novel information measure that can measure the discernibility power of crisp as well as fuzzy equivalence relations. They further proposed a general definition of significance for nominal, numerical, and fuzzy attributes [20]. The method proposed by Wu [50] handles feature reduction in incomplete information systems where some of the attributes are non-deterministic (i.e., missing or partially known). More complete reviews of rough set-based feature selection techniques are available elsewhere [45,47].

Studying the importance of feature values in a dataset can provide a different perspective on the data. Contrary to feature selection, there has been very little research on computing the importance of feature values [42]. This problem is very different from identifying significant features because even an insignificant feature may have some important values, whereas some values of a significant feature may not be very important. In the next section, we propose a rule-based method for identifying a three-factor structure in customer satisfaction data.

3. The proposed rule-based method

In this section, we introduce a rule-based method to identify different types of features in customer satisfaction datasets. The inputs for these rules are created by using a feature selection method [1]. The method computes the significance of features by using feature value distributions over different classes.

3.1. Method to compute the significance of feature values

Ahmad and Dey [1] proposed an efficient method to compute the significance of categorical features in a database by using conditional probabilities. The significance measure proposed in their study is based on the rationale that each feature value has different associations with different class subsets and that the subset for which the association reaches a maximum is defined as the support set of the feature value. Additionally, a significant feature is likely to have different support sets for different feature values, while this may not be so for an insignificant feature. Given a feature value, its relative frequency in different classes is used to measure the feature value-to-class association. We introduce a set of notations and explain the computations involved as follows:

(a) Let U be the collection of pre-classified data elements. Let the elements of U be members of m different classes, where m classes are denoted by $1, 2, 3 \ldots, m$. Let J represent the set of all class labels, i.e., $J = \{1, 2, 3, \ldots, m\}$.
(b) Let $A_1, A_2, \ldots, A_g$ be the features describing the elements of this dataset.
(c) Let A_i^r denote the rth feature value of the ith feature A_i.
(d) $\sim A_i^r$ is a generic notation used to denote any value of A_i that is not equal to A_i^r.
(e) Let w be a proper subset of J, w is a subset of class labels. $(\sim w)$ denotes the subset of labels that are not contained in w. Thus $\sim w = J - w$.
(f) Let $p(w/A_i^r)$ denote the probability that the elements of U with ith feature value equal to A_i^r belong to a class contained in w. This probability can be computed from U by using frequency counts.
(g) Let $p(\sim w/\sim A_i^r)$ denote the probability that elements not having the ith feature value equal to (A_i^r) do not belong to a class contained in w. In other words, given a feature value A_i^r, $p(\sim w/\sim A_i^r)$ denotes the probability that the elements not having the feature value A_i^r belong to the set $\sim w$. This can also be computed from U by using frequency counts.

Ahmad and Dey [1] observed that if a feature value is significant, both $p(w/A_i^r)$ and $p(\sim w/\sim A_i^r)$ will be high. This behavior implies that objects with value A_i^r for the ith feature A_i, and objects with values indicated by $\sim A_i^r$, would belong to different groups of complementary classes. Because w denotes a proper subset of classes, w can take $2^m - 1$ different values, corresponding to which the quantity $p(w/A_i^r) + p(\sim w/\sim A_i^r)$ can assume $2^m - 1$ different values. The quantity $max(p(w/A_i^r) + p(\sim w/\sim A_i^r) - 1)$, which corresponds to the maximum of the $2^m - 1$ values, is termed as the discriminating power of the feature value A_i^r. The subset w that yields the maximum value is denoted by w_i^r.

Definition 3.1.1. The subset $w = w_i^r$, which maximizes the term $(p(w/A_i^r) + p(\sim w/\sim A_i^r) - 1)$, is termed the *support set* for the value A_i^r. Because w_i^r yields the maximum value for the above quantity, the *support set* is said to have the strongest association among all the possible proper subsets of J to the value A_i^r.

Definition 3.1.2. The quantity $\vartheta_i^r = p(w_i^r/A_i^r) + p(\sim w_i^r/\sim A_i^r) - 1$ is defined as the discriminating power of feature value A_i^r, where w_i^r is the support set for the value A_i^r.

A feature value is important if it has a large discriminating power. The value of the discriminating power of a feature value A_i^r lies between 0 and 1.

Support sets and discriminating powers provide a good visualization for the distinguishing characteristics of the different feature values. Based on previous discussions and analyses, the following observations can be made about the importance of feature values within the dataset.

(a) Feature values with similar properties have similar support sets, and feature values with different properties have different support sets.
(b) If two feature values have similar support sets, then the feature value with greater discriminating power is more significant.
(c) Feature values with insignificant discriminating powers have no effect on the dependent variable.

3.1.1. The algorithm to compute support sets and discriminating powers

Ahmad and Dey [1] proposed an efficient algorithm to compute feature value support sets and discriminating powers. This algorithm is linear with respect to the number of data points and uses the conditional probabilities of classes co-occurring with various feature values. The significance of a feature is computed as the mean of the discriminating powers of all its values. The simplicity of the algorithm is derived from the fact that for each value, it checks its association with a particular class only once and not for each occurrence of it in the power-set of classes. This algorithm is presented in Fig. 1; however, the computational process to find a support set and a discriminating power is explained by running through an example dataset, which is presented in Table 1. This toy dataset contains nine elements, described by two categorical features, belonging to three different classes. For feature 1, the possible values are A, B and C; for feature 2, the possible values are a, b and c. The computations for value A of feature 1 are demonstrated as follows:

Step 1 Compute conditional probabilities $p(\gamma/v)$, where γ is a class and v is a feature value. In this case, γ varies from 1 to 3, while v is A; hence,

$$p(1/A) = 3/4, \quad p(1/\sim A) = 1/5.$$
$$p(2/A) = 1/4, \quad p(2/\sim A) = 2/5.$$
$$p(3/A) = 0, \quad p(3/\sim A) = 2/5.$$

Step 2 Find the support sets for the feature values.
Because $p(1/A) > p(1/{\sim}A)$, add 1 to the support set of A.
Since $p(2/A) < p(2/{\sim}A)$, do not add 2 to support set of A.
Since $p(3/A) < p(3/{\sim}A)$, do not add 3 to support set of A.

Thus the support set of the feature value A is {1}. Hence, the discriminating power of the feature value A is given by

$$\begin{aligned} p(1/A) + p(\sim 1/\sim A) - 1 &= p(1/A) + (p(2/\sim A) + p(3/\sim A)) - 1. \\ &= 3/4 + (2/5 + 2/5) - 1. \\ &= 11/20. \end{aligned}$$

3.2. The central idea of the proposed method

We propose a rule-based method for identifying different types of features in customer satisfaction datasets. These rules are based on support sets and the discriminating powers of feature values.

(a) *Basic features* – Customers take these basic features for granted. Below the threshold value, these feature values contribute to customer dissatisfaction, whereas above the threshold value, they contribute very little towards customer satisfaction. For these types of features, low feature values (i.e., below the threshold value) have similar support sets (consisting of customer dissatisfaction values) and similar discriminating powers, whereas high feature values (i.e.,

```
Algorithm_ discriminating_power
/*  It computes the support set and the discriminating power of a feature value A_i^r */
Begin
        ϑ_i^r = 0 ; /* discriminating power initialized to 0 */
        w_i^r = φ ; /* Support set initialized to NULL */
        For (t = 1; t ≤ m; t++)   /* m number of classes*/
          {
            If  p(t / A_i^r) > p(t/~A_i^r)) / * t occurs more frequently with A_i^r than
                {                                 with ~ A_i^r.*/
                  1-   Add t  to w_i^r;     /* t is added to the support set.*/
                  2-    ϑ_i^r =   ϑ_i^r + p(t/ A_i^r);
                }
            Else
                  {   ϑ_i^r =   ϑ_i^r + p(t /~A_i^r);   }
          }
        endfor
            ϑ_i^r =  ϑ_i^r - 1;
End
```

Fig. 1. The algorithm to compute the support set and the discriminating power of a feature value [1].

Table 1
An example dataset.

Feature 1	Feature 2	Class
A	*a*	1
A	*a*	1
A	*b*	1
A	*b*	2
B	*c*	2
B	*c*	2
C	*c*	1
C	*c*	3
C	*c*	3

above the threshold value) have similar support sets (consisting of customer satisfaction values) and similar discriminating powers. In other words, if we find that for a feature, it's values have only two types of support sets, one with customer dissatisfaction values and another with customer satisfaction values, we can predict that the feature is a basic feature. *The rule for basic features can be written as follows:*
If feature values have only two types of support sets (one with customer dissatisfaction values and another with customer satisfaction values),
then the feature is a basic feature.

(b) *Performance features* – These features have both positive and negative effects on customer satisfaction. As the values of a performance feature increase, there is a gradual shift in their contribution from strong customer dissatisfaction to strong customer satisfaction. For a performance feature, different feature values (e.g., strong dissatisfaction values, weak dissatisfaction values, weak satisfaction values, and strong satisfaction values) have different support sets. In other words, if feature values have different support sets, we can predict that the feature is a performance feature. *The rule for performance features can be written as follows:*
If feature values have different support sets that change gradually from strong dissatisfaction values to strong satisfaction values,
then the feature is a performance feature.

(c) *Excitement features* – For small feature values, these features do not contribute to customer dissatisfaction or customer satisfaction; however, for large feature values, excitement features contribute only positively towards customer satisfaction. For an excitement feature, some of the feature values have similar support sets that are mixed in nature (i.e., consisting of customer dissatisfaction and customer satisfaction values) and these feature values possess low discriminating powers, whereas the remaining feature values have support sets consisting of strong customer satisfaction values. *The rule for the excitement features can be written as follows:*
If some of the feature values have similar and mixed support sets with low discriminating powers *and* the remaining feature values have support sets with strong customer satisfaction values,
then the feature is an excitement feature.

(d) *Random features* – These features contribute neither to customer satisfaction nor dissatisfaction. Hence, all the feature values of these features have similar support sets consisting of customer satisfaction and dissatisfaction values. All the feature values of these features have low discriminating powers. The rule of the random feature can be written as follows:
If all the feature values have similar support sets consisting of customer satisfaction and dissatisfaction values *and* all the feature values have low discriminating powers,
then the feature is a random feature.

As the proposed rules use the feature selection method [1], which is linear with respect to the number of data points, *our method is more computationally efficient than the method suggested by Sikonja and Vanhoof* [42] (their method calculates the k-nearest neighbors of each member of training points, making the process slow (see Section 1)). Their method works only for ordinal features, whereas *this proposed method works for categorical features;* however, it is easy to convert numeric features to the categorical features by the discretization process, and hence, this method is useful for all types of datasets.

In the next section, we apply our proposed method to study customer satisfaction datasets within a three-factor theoretical framework.

4. Experiments

In this section, we study the effectiveness of our proposed method on a synthetic dataset with known characteristics and also on a real housing dataset.

Sikonja and Vanhoof [42] presented a synthetic dataset that simulates customer satisfaction. This dataset contains three significant features, one *basic* (B), one *performance* (P), and one *excitement* (E), where the roles of B, P, and E are similar to those explained in Section 1. This dataset also contains one *random* feature (R) and a dependent variable C that denotes *customer satisfaction*. In addition, $b(B)$ denotes the effect of the basic feature on customer satisfaction, $p(P)$ denotes the effect of the performance feature on customer satisfaction, and $e(E)$ defines the effect of the performance feature on customer satisfaction. $r(R)$ defines the effect of the random feature on the customer satisfaction value which is set to 0. The values of all the other features are randomly generated integer values from 1 to 5. Customer satisfaction values were generated by using the following equation,

$$C = b(B) + p(P) + e(E) + r(R) \tag{5}$$

where a higher value of C indicates a higher degree of satisfaction.

The functions b, p and e are defined as follows:

$$\begin{aligned} b(B) &= -1, \quad \text{when } B <= 3 \\ &= 1, \quad \text{when } B > 3 \\ p(P) &= P - 2. \\ e(E) &= 0, \quad \text{when } E <= 4 \\ &= 2, \quad \text{when } E = 5 \end{aligned}$$

These relationships imply that if the basic feature B has a value less than or equal to 3, then it has a negative impact on satisfaction, whereas feature values 4 and 5 have a positive impact on satisfaction. It may be noted that the impact, positive or negative, is independent of the actual values of B and only depends on whether or not it is greater than, less than or equal to 3. The effect of the performance feature P on satisfaction grows positively as P increases; therefore, the relationship with customer satisfaction is linear. For lower values of P, dissatisfaction exists; however, after P crosses a threshold value (2 in this case), it has a positive effect on customer satisfaction, and the effect grows linearly as the value of P increases. The excitement feature E contributes to customer satisfaction only when it occurs with a value 5. Random feature R does not affect customer satisfaction. For more information on this synthetic dataset, one can refer to [42]. Using the function given in Eq. (5), we generated 1000 data points, with C varying from −2 to 6. Each feature has values in the range of 1–5. This dataset was then used to compute support sets and discriminating power values of different feature values by using the conditional probability-based approach [1]. In the next subsection, we show that by using our proposed rules (Section 3.2), one can capture the above-mentioned behavioral aspects of the features.

Table 2
Support sets of different values of the basic feature *B*.

Feature value	Support set
1	{−2, −1, 0, 1}
2	{−2, −1, 0, 1}
3	{−2, −1, 0, 2}
4	{2, 3, 4, 5, 6}
5	{3, 4, 5, 6}

4.1. Analysis of results

As the proposed approach does not use the feature value ordering to compute support sets and discrimination powers of feature values, the data generated above were treated as a categorical dataset. The customer satisfaction values were also treated as categorical values, with negative values indicating dissatisfaction, positive values indicating satisfaction, and 0 indicating neither of the two. Table 2 presents the support sets obtained for the different values of basic feature *B*. It can be observed from Table 2 that the feature values 1, 2 and 3 for *B* have almost the same support sets, and most of the classes that belong to these support sets are either negative or zero. For the feature values 4 and 5, we obtained positive values of *C* in the support set. This result suggests that the feature values 4 and 5 are more likely to contribute to customer satisfaction, while the values 1, 2 and 3 are more likely to cause customer dissatisfaction. By computing the discriminating powers of the two values 4 and 5, the value that contributes more towards satisfaction can be determined. It was found that they had almost identical discriminating powers (0.28 for the feature value 4 and 0.30 for the feature value 5). This similarity indicates that they affect customer satisfaction in almost identical ways. It can be seen that feature values for this feature have two types of support sets, which is similar to the conditions for a basic feature (Section 3). Hence, one can predict that this is a basic feature, which matches the known characteristic of the feature.

Sikonja and Vanhoof [42] further computed the effect of feature values in terms of upward and downward reinforcement. They report that the feature value 4 has a strong upward reinforcement value and that the feature value 3 has a strong downward reinforcement value. It is also possible to extract these characteristics by using the support sets. Support sets for different feature values of *B* are shown in Fig. 2. It may be observed that as we move from the feature value 3 to the feature value 4, there is a sudden jump in support set values, from negative to positive classes. This jump shows that the feature value 4 contributes strongly to the upward movement of customer satisfaction. Similarly, when we move from feature value 4 to 3, there is a sudden drop in support set values from positive to negative. This drop suggests that the feature value 3 contributes strongly to the downward movement of customer satisfaction. Thus, our findings are consistent with the findings of Sikonja and Vanhoof [42] for these feature values.

Table 3 shows the support sets for various values of the performance feature *P* computed using our approach. It is observed that the feature value 1 has support set (0, −2), the feature value 2 has support set (−1, 1), the feature value 3 has support set (0, 2), the feature value 4 has support set (1, 3, 5) and the feature value 5 has support set (2, 4, 6). All the feature values have different support sets, and these support sets change gradually from strong dissatisfaction values to strong satisfaction values. This characteristic is similar to the conditions for a performance feature. Hence, based on our proposed rule-based method, this feature can be identified as a performance feature, and thus, our prediction matches the characteristic of the performance feature.

For the performance feature *P*, Sikonja and Vanhoof [38] report that the strongest upward reinforcement is observed for the feature value 3. Upward reinforcements for feature values 4 and 5 are also strong. The feature value 2 is reported to have

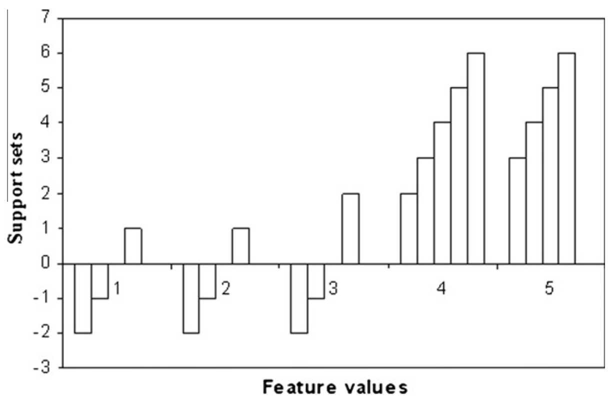

Fig. 2. Feature value vs. support sets graph for the basic feature *B*.

Table 3
Support sets of different values of the performance feature *P*.

Feature value	Support set
1	{0, −2}
2	{1, −1}
3	{1, 2}
4	{1, 3, 5}
5	{2, 4, 6}

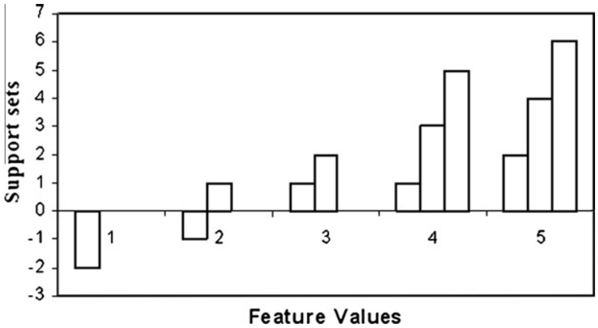

Fig. 3. Feature value vs. support sets graph for the performance feature *P*.

the strongest downward reinforcement, and feature values 3 and 4 also show strong downward reinforcements. Support sets of different feature values are also presented in Fig. 3. As we move from the feature value 2 to 3, the behavior of the support set changes from a mixed support set to all positive values for customer satisfaction, which justifies the strongest upward reinforcement for the feature value 3. As we move to the feature value 4 from 3, it is observed that the behavior of the support sets remains identical, i.e., both values are indicative of customer satisfaction only. However, the actual values in these support sets shift to the more positive side, thus indicating a stronger upward reinforcement power for the feature value 4. Similar observations can be made for the feature value 5, which also shows a strong upward reinforcement. As we move in the opposite direction, from the feature value 5 to 4, the class values in the support set decrease. This decrement explains the downward reinforcement nature for the feature value 4. As we move from the feature value 3 to 2, the behavior of the support set changes from all positive classes to mixed classes, which indicates that the feature value 2 has the strongest downward reinforcement. All the observations are in agreement with those of Sikonja and Vanhoof [42].

Table 4 presents support sets for different values of the excitement feature *E*. For the feature values 1 to 4, the support sets are almost similar, but as the values change from 4 to 5, the character of the support set changes. Moreover, the discriminating powers of all the feature values from 1 to 4 are very low (less than 0.1). Most of the feature values of this feature have similar and mixed support sets, and the discriminating powers of these features are low. However, the feature value 5 has a support set consisting of strong satisfaction values. These conditions are similar to the rules for an excitement feature. Hence, based on our proposed rule-based method, this feature is identified as an excitement feature, which matches the characteristics of this feature.

Fig. 4 shows that there is a sudden jump in satisfaction values in the support set for the feature value 5, indicating that the value 5 contributes strongly to customer satisfaction. Conversely, a move from the feature value 5 (the support set with positive values) to 4 (the support set with mixed values) exhibits a strong downward force, showing that the feature value 4 contributes strongly to the lower values of customer satisfaction. These observations are consistent with the findings of Sikonja and Vanhoof [42]. Table 5 and Fig. 5 show the support sets for values of the random feature *R*. All the feature values of *R* have similar support sets consisting of both positive and negative values of customer satisfaction. The discriminating powers of all the feature values are very low, ranging from 0.01 to 0.05. These observations follow the conditions for a random feature. Hence, on the basis of our rule-based method, one may predict correctly the category of this random feature. As these

Table 4
Support sets of different values of the excitement feature *E*.

Feature value	Support set
1	{−2, −1, 0, 1}
2	{−2, −1, 0, 1, 2}
3	{−2, −1, 0, 2}
4	{−2, −1, 0, 1}
5	{3, 4, 5, 6}

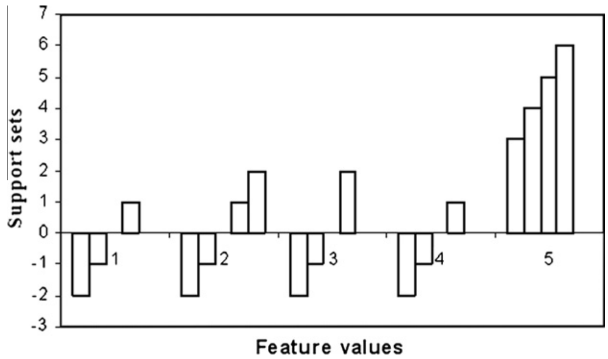

Fig. 4. Feature value vs. support sets graph for the excitement feature *E*.

Table 5
Support sets of different values of the random feature *R*.

Feature value	Support set
1	{−2, −1, 1, 5, 6}
2	{−2, −1, 2, 4}
3	{−2, −1, 0, 3, 4}
4	{−2, −1, 1, 4}
5	{−2, −1, 0, 5}

feature values have very low discriminating powers, these values have negligible effect on customer satisfaction. Sikonja and Vanhoof [42] report similar findings of very small upward reinforcement and downward reinforcement for all the feature values.

This analysis suggests that by using our rule-based method, we can predict the behavior of different features in a customer satisfaction dataset. In the next subsection, we present our results on a real housing dataset.

4.2. Experiments on a real housing dataset

We report the performance of our approach using a Boston housing dataset. These data are taken from UCI's machine learning repository (www.ics.uci.edu/∼mlearn/MLRepository.html). The dataset has 1 binary feature and 12 continuous features. Information about these features is presented in Table 6. The task associated to this dataset was to predict the prices of houses. Homburg et al. suggested [18] that customer satisfaction has a strong positive association to the price a customer is willing to pay for a house. Thus, the price can be treated as a measure of customer satisfaction. In this case, there is no explicit measure of dissatisfaction. The low price of a house may be assumed to be due to dissatisfaction. All continuous features, including price, are converted into integer-valued features (feature values 1–5) by using equal-width discretization.

Support sets for various feature values are presented in Table 7. One can predict the nature of these features on the basis of the rules proposed in Section 3.2. On the basis of these rules, it was concluded that the *per-capita crime rate by town, lower socioeconomic status of the population* and *proportion of African–Americans by town*, are *basic features* (Table 8). Interestingly, all these features are generic social features associated with a neighborhood, whereas the rest of the features are related to individual house characteristics (such as *full-value property-tax*) or facilities proximal to the house (such as *pupil-teacher ratio*

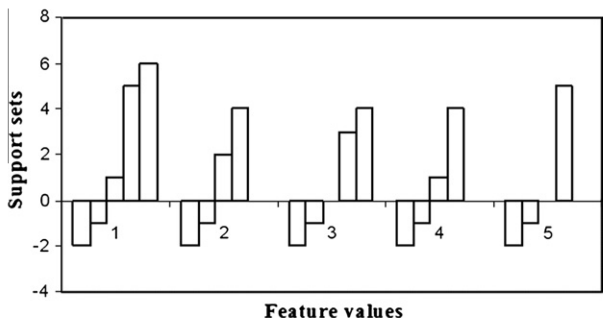

Fig. 5. Feature values vs. support sets graph for the random feature *R*.

Table 6
The housing dataset features information.

No.	Feature description
1	Per capita crime rate by town
2	Proportion of residential land zoned for lots over 25,000 sq.ft.
3	Proportion of non-retail business acres per town
4	Charles River dummy variable (= 2 if tract bounds river, and 1 otherwise)
5	Nitric-oxide concentration (parts per 10 million)
6	Average number of rooms per dwelling
7	Proportion of owner-occupied units built prior to 1940
8	Weighted distances to five Boston employment centers
9	Index of accessibility to radial highways
10	Full-value property tax rate per $10,000
11	Pupil-teacher ratio by town
12	1000(AA – 0.63)^2 where AA is the proportion of Afro-American by town
13	% Lower socioeconomic status of the population

Table 7
Support sets of different feature values for the housing dataset. Each row presents a feature and each column presents a feature value.

Feature\Feature values	1	2	3	4	5
1	{2, 3, 4, 5}	{1}	{1}	{1}	{1}
2	{1, 2}	{3, 4, 5}	{3, 4}	{3, 4}	{3, 4, 5}
3	{3, 4, 5}	{2, 3}	{2, 3}	{1, 2}	{1, 2}
4	{1, 2, 3}	{4, 5}			
5	{3, 4, 5}	{2, 3}	{2, 1}	{1}	{1, 2}
6	{1}	{1, 2}	{2, 3}	{3, 4, 5}	{4, 5}
7	{3, 4, 5}	{3, 4, 5}	{2, 3, 4}	{2, 3, 5}	{1, 2}
8	{1, 2, 5}	{3, 4}	{2, 3, 4}	{3, 4}	{2, 3}
9	{2, 3}	{2, 3}	{3, 4}	{4, 5}	{1}
10	{3, 4, 5}	{2, 3}	{2, 4, 5}		{1}
11	{4, 5}	{2, 3, 4}	{2, 3, 4}	{2, 3}	{1, 2}
12	{1}	{1}	{1, 3}	{1, 2}	{2, 3, 4, 5}
13	{3, 4, 5}	{2}	{1, 2}	{1}	{1}

by town). Social features play a very important role in house valuation. A study [30] has suggested that houses are highly discounted in high crime areas and that neighborhood poverty also reduces the price of a house. This effect can be seen as a "peer group effect" [12] as people belonging to same economic class tend to live in the same localities.

The proportion of non-retail business acres per town, nitric-oxides concentration, the average number of rooms per dwelling and *full-value property-tax rate* are identified as performance features (Table 8). Other than the feature *proportion of non-retail business acres per town*, all the features are related to individual house characteristics. In general, people tend to pay more for bigger houses, and a higher tax rate has an adverse effect on the house price.

The features *proportion of residential land zoned for lots over 25,000 sq.ft., proportion of owner-occupied units built prior to 1940, index of accessibility to radial highways* and *pupil-teacher ratio by town* (Table 8) are found to be excitement features according to our model. These features may be viewed as descriptors of facilities proximal to a house.

Generally, in any real dataset, some features show mixed patterns and cannot be characterized completely into one category of the three-factor approach [42]. A similar behavior is also observed with some features in this dataset. For example, for the feature *index of accessibility to radial highways*, there is initially a steady increase of house prices with a rise in feature values. However, for very large values, there is a sharp decrease in the price.

Table 8
Characteristics of different features of the housing dataset obtained by our proposed method.

The type of features	Features	Comments
Basic	1, 12, 13	Values should be very high for better price***
Performance	3, 5, 6, 10*	Low values relates to low prices. Steadily increase in prices with values***
Excitement	2, 7, 9**, 11	These values do not contribute to low prices***
Random	8	No relationship between feature values and prices

* Generally increasing, the support set of the feature value 3 does not follow the trend.
** The very large value leads to the sharp decrease in the price.
*** Low values and high values are different for different features as some features have negative correlations, whereas some features have positive correlations. For example, for the crime feature, the feature value 1 presents a large positive effect. Attribute 4 has only 2 values, it is not included in this categorization.

Our results suggest that the physical nature is different for different types of features: *basic* (neighborhood), *performance* (individual house characteristics) and *excitement* (facilities proximal to the house). More insights into the causes and effects of these relationships may be elucidated by domain experts. However, the proposed model characterized these features accurately, even without any domain knowledge, thus showing the effectiveness of our proposed rule-based method.

Our method could be utilized to assess the accuracy of housing prices based on the nature of different feature values. For example, on the one hand, one may concentrate only on low values of basic features to find a house that is not highly priced. On the other hand, while inquiring about a high-valued house, one should evaluate performance features and excitement features because the high values of these features would indicate a good investment.

5. Conclusion

It has been established that in market research, knowing which feature values contribute to customer satisfaction and customer dissatisfaction and in what way they influence customer behavior is crucial. Market managers and practitioners will be able to make better decisions if they have a better understanding of customer satisfaction data. In this paper, we have presented a rule-based method to analyse the behavior of features in customer satisfaction. The inputs for these rules were created by using a probabilistic feature selection method that computes the support set and discriminating power for each feature value. In general, this method can help business managers identify the critical factors in customer satisfaction survey data so that business managers can make data-informed decisions about how to best allocate limited resources to attain the highest degree of customer satisfaction. Our proposed method is computationally efficient and can handle all types of features. The proposed approach was applied on a synthetic customer satisfaction dataset for which the feature characteristics had previously been established. The results suggest that our proposed method successfully extracted these characteristics from the dataset. We also performed experiments on a real housing dataset. The characteristics of the data obtained by our proposed method match the domain knowledge, which shows the effectiveness of our proposed method on the real housing dataset. The datasets used in this paper were non-categorical but were treated as categorical; in the future we will apply this method to purely categorical customer satisfaction datasets. We will also apply our method to study customer satisfaction for different products and services.

References

[1] A. Ahmad, L. Dey, A feature selection technique for classificatory analysis, Pattern Recognition Letters 26 (1) (2005) 43–56.
[2] R. Agrawal, R. Srikant, Mining sequential patterns, in: 11th International Conference on Data Engineering (ICDE'95), 1995, pp. 3–14.
[3] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th international conference on very large databases, Santiago, Chili, 1994, pp. 487–499.
[4] E.W. Anderson, V. Mittal, Strengthening the satisfaction-profit chain, Journal of Service Research 3 (2) (2000) 107–120.
[5] L.L. Berry, Relationship marketing, in: L.L. Berry, G. Shostack, G. Upah, (Eds.), Proceedings of Emerging Perspectives in Service Marketing, Chicago, 1983, pp. 25–28.
[6] M.J.A. Berry, G.S. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Support, John Wiley and Sons, 1997.
[7] M. Beynon, Reducts within the variable precision rough sets model: a further investigation, European Journal of Operational Research 134 (2001) 592–605.
[8] L.A. Blum, P. Langley, Selection of relevant features and examples in machine learning, Artificial Intelligence 97 (1997) 245–271.
[9] H.H. Bock, The Classical Data Situation, Analysis of Symbolic Data, Springer, 2002. pp. 139–152.
[10] R.N. Bolton, H.D. James, A multi-stage model of customer's assessments of service quality and value, Journal of Consumer Research 17 (March) (1991) 375–384.
[11] A.C. Burns, R.F. Bush, Marketing Research, Prentice Hall, 2005.
[12] P. Cheshire, S. Sheppard, Introduction to feature: the price of access to better neighbourhoods, The Economic Journal 114 (November) (2004) F391–F396.
[13] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, The KDD process for extracting useful knowledge from volumes of data, Communications of the ACM 39 (11) (1996) 27–34.
[14] J. Fuller, K. Matzler, Customer delight and market segmentation: an application of the three-factor theory of customer satisfaction on life style groups, Tourism Management 29 (1) (2008) 116–126.
[15] B. Galitsky, J.L. de la Rosa, Concept-based learning of human behaviour for customer relationship management, Information Sciences 181 (2011) 2016–2035.
[16] B. Ganter, R. Wille, Formal Concept Analysis, Mathematical Foundations, Springer, 1999.
[17] R. Hanson, Determining attributes importance, Quirk's Marketing Research Review 6 (October) (1992) 16–18.
[18] C. Homburg, N. Koschate, W.D. Hoyer, Do satisfied customers really pay more? a study of the relationship between customer satisfaction and willingness to pay, Journal of Marketing 69 (2) (2005) 84–96.
[19] Q. Hu, Z. Xie, D. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern Recognition 40 (12) (2007) 3509–3521.
[20] Q. Hu, D. Yu, Z. Xie, Information-preserving hybrid data reduction based on fuzzy-rough techniques, Pattern recognition letters 27 (5) (2006) 414–423.
[21] J.A. Hyung, Evaluating customer aid functions of online stores with agent-based models of customer behaviour and evolution strategy, Information Sciences 180 (9) (2010) 1555–1570.
[22] R. Johnston, The determinant of service quality: satisfiers and dissatisfies, International Journal of Service Industry management 6 (5) (1995) 53–71.
[23] G.H. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selection problem, in: Proceedings of 11th International Conference on Machine Learning, Morgan Kaufman, New Brunswick, NJ, 1994, pp. 121–129.
[24] N. Kano, S. Nobuhiku, T. Fumio, T. Shinichi, Attractive quality and must-be quality, The Journal of the Japanese Society for Quality Control (1984) 39–48.
[25] K. Kira, L. Rendell, A practical approach to feature selection, in: Proceedings of Ninth International Conference on Machine Learning, Aberdeen, Scotland, 1992, pp. 249–256.
[26] I. Kononenko, Estimating features: analysis and extensions of RELIEF, in: Proceedings of European Conference on Machine Learning, 1994, pp. 171–182.

[27] W. Lee, B. Shih, L. Tu, The application of Kano's model for improving web-based learning performance, Frontiers in Education (FIE 2002) (2002). pp. T3E-27–T3E-32.
[28] K. Li, Y.S. Liu, Rough set based attribute reduction approach in data mining, in: Proceedings of the First International Conference on Machine Learning and Cybernetics, vol. 1, 2002, pp. 60–63.
[29] J.J.H. Liou, G.H. Tzeng, A dominance-based rough set approach to customer behaviour in the airline market, Information Sciences 180 (11) (2010) 2230–2238.
[30] A.K. Lynch, D.W. Rasmussenz, Measuring the impact of crime on house prices, Applied Economics 33 (2001) 1981–1989.
[31] K. Matzler, F. Bailom, H.H. Hinterhuber, B. Renzl, J. Pichler, The asymmetric relationship between attribute-level performance and overall customer satisfaction: a reconsideration of the importance–performance analysis, Industrial Marketing Management 33 (4) (2004) 271–277.
[32] V. Mittal, W.A. Kamakura, Satisfaction, repurchase intent, and repurchase behaviour: investigating the moderate effect of customer characteristics, Journal of Marketing Research 38 (February) (2001) 131–142.
[33] P.M. Narendra, K. Fukunaga, A branch and bound algorithm for feature subset selection, IEEE Transactions on Computer C-26 (9) (1977) 917–922.
[34] E. Nassiri-Mofakham, M.A. Nematbakhsh, A. Baraani-Dastjerdi, N. Ghasem-Aghaee, Electronic promotion to new customers using mkNN learning, Information Sciences 179 (3) (2009) 248–266.
[35] G. Pandula, B. Busacca, The asymmetric impact of price-attribute performance on overall price evaluation, International Journal of Service Industry Management 16 (1) (2005) 28–54.
[36] M. Paulssen, A. Sommerfeld, Modeling the nonlinear relationship between satisfaction and loyalty with structural equation models, in: Proceedings of the 29th Annual Conference From Data and Information Analysis to Knowledge Engineering, Gesellschaft für Klassifikation e.V., University of Magdeburg, 2005, pp. 574–581.
[37] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341–356.
[38] Z. Pawlak, A. Skowron, Rudiments of rough sets, Information Sciences 177 (1) (2007) 3–27.
[39] Z. Pawlak, A. Skowron, A rough sets: some extensions, Information Sciences 177 (1) (2007) 28–40.
[40] J.R. Quinlan, Induction of decision trees, Machine Learning 1 (1986) 81–106.
[41] H.W. Shin, S.Y. Sohn, Segmentation of stock trading customers according to potential value, Expert systems with applications 27 (1) (2004) 27–33.
[42] M.R. Sikonja, K. Vanhoof, Evaluation of ordinal features at value level, Data Mining and Knowledge Discovery 14 (2) (2007) 225–243.
[43] R. Srikant, R. Agrawal, Mining generalized association rules. VLDB'95, 1995, pp. 407–419.
[44] B. Stauss, B. Hentshel, Feature-based versus incident-based measurement of service quality: result of an empirical study in German car service industry, in: P. Kunst, J. Lemmink, (Eds.), Quality Management in Services, Van Gorcum, Assen, 1992, pp. 59–78.
[45] R.W. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Letters 24 (6) (2003) 833–849.
[46] H. Takeuchi, L.V. Subramaniam, T. Nasukawa, S. Roy, Getting insights from the voices of customers: conversation mining at a contact centre, Information Sciences 179 (11) (2009) 1584–1591.
[47] K. Thangavel, A. Pethalakshmi, Dimensionality reduction based on rough set theory: a review, Applied Soft Computing 9 (1) (2009) 1–12.
[48] R. Weber, Customer segmentation for banks and insurance groups with fuzzy clustering techniques, in: J.F. Baldwin, (Ed.), Proceedings of Fuzzy Logic, Wiley, New York, NY., 1996, pp. 187–196.
[49] D.R. Wittink, L.R. Bayer, The measurement imperative, Marketing Research 6 (1994) 14–23.
[50] W.Z. Wu, Attribute reduction based on evidence theory in incomplete decision systems, Information Sciences 178 (5) (2008) 1355–1371.