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]