Available online at www.sciencedirect.com

ScienceDirect

Procedia Economics and Finance 17 (2014) 3 – 9

Procedia Economics and Finance

www.elsevier.com/locate/procedia

Innovation and Society 2013 Conference, IES 2013

Bayesian Network Applications to Customer Surveys and InfoQ

Federica Cugnata[a], Ron Kenett[b], Silvia Salini[c]*

[a] Department of Economics and Statistics 'Cognetti De Martiis', University of turin, Italy
[b] KPA Ltd., Raanana, Israel Center for Risk Engineering, NYU-Poly, NY, USA
[c] Department of Economics, Management and qunatitative Methods, University of Milan, Italy

Abstract

Modelling relationships between variables has been a major challenge for statisticians in a wide range of application areas. In conducting customer satisfaction surveys, one main objective, is to identify the drivers to overall satisfaction (or dissatisfaction) in order to initiate proactive actions for containing problems and/or improving customer satisfaction. Bayesian Networks (BN) combine graphical analysis with Bayesian analysis to represent relations linking measured and target variables. Such graphical maps are used for diagnostic and predictive analytics. This paper is about the use of BN in the analysis of customer survey data. We propose an approach to sensitivity analysis for identifying the drivers of overall satisfaction. We also address the problem of selection of robust networks. Moreover, we show how such an analysis generates high information quality (InfoQ) and can be effectively combined with an integrated analysis considering various models.

Selection and peer-review under responsibility of the Organizing Committee of IES 2013.

Keywords: Customer Satisfaction; Importance-performance analysis

1. Introduction

Customer satisfaction studies deal with customers, consumers and user satisfaction from a product or service. The topic was initially developed in marketing theory and applications. The Business Dictionary defines customer satisfaction as "The degree of satisfaction provided by the goods or services of a company as measured by the number of repeat customers." (Kenett and Salini, Chapter 1, 2012). With such a definition, customer satisfaction

* Corresponding author. Tel.: +97297408442 ; fax: +97297408443.
E-mail address: ron@kpa-group.com

Selection and peer-review under responsibility of the Organizing Committee of IES 2013.
doi:10.1016/S2212-5671(14)00871-5

seems to be an objective and easily measured quantity. However, unlike variables such as type of product purchased or geographical location, customer satisfaction is not necessarily observed directly. Typically, in a social science context, analysis of such measures is performed indirectly by employing proxy variables. Unobserved variables are referred to as latent variables, whilst proxy variables are known as observed variables. In many cases, the latent variables are very complex and the choice of suitable proxy variables is not immediately obvious. For example, in order to assess customer satisfaction from an airline service, it is necessary to identify attributes that characterize this type of service. A general framework for assessing airlines includes attributes such as: on board service, timeliness, responsiveness of personnel, airplane seats and other tangible service characteristics. In general, some attributes are objective, related to the service's technical-specific characteristics and others are subjective, dealing with behaviors, feelings and psychological benefits Statistical analysis is a science that relies on a transformation of the content domain space into an analytic space with dimensions that lend themselves to quantitative analysis. Self-administered surveys use structured questioning to map out perceptions and satisfaction level, using observations from a population frame, into data that can be statistically analyzed.

The importance that users or customers attach to various services and products is a vital part of customer satisfaction surveys, as much as the measure of the quality and the satisfaction.
In some cases, the level of importance is asked explicitly in the survey questionnaire; in other cases, it is derived using statistical models.

The survey process consists of four main stages: 1) Planning, 2) Collecting, 3) Analyzing and 4) Presenting. Modern surveys are conducted in a wide variety of techniques including phone interviews, self-reported paper questionnaires, email questionnaires, internet-based surveys, SMS-based surveys, face to face interviews, videoconferencing etc.

Eventually, the survey is providing information to decision makers. In this paper we focus on one specific aspect of the information derived from satisfaction surveys which is related to the link between level of satisfaction and importance attributed by customers to survey items. The analysis we propose is designed to increase the quality of information when the survey goals are to identify decision drivers affecting customer decisions. The next section briefly summarizes the various dimensions of information quality and the concept of InfoQ developed by Kenett and Shmueli (2013). The following section presents an analysis and sensitivity assessment of satisfaction drivers using Bayesian Network models and the final section concludes with a summary and a description of further areas of research.

2. Information Quality of Customer Surveys

Information quality (InfoQ) defined as the potential of a dataset to achieve a specific (scientific or practical) goal using a given empirical analysis method (Kenett and Shmueli, 2013). In assessing InfoQ one first needs to describe a specific research study with four components: i) a specific analysis goal (g), ii) the available dataset (X), iii) the method or model that was used (f) and iv) a utility measure (U). As a generic concept, we define , i.e. the derived utility from an application of a model to a certain data set, given the research goals. In our case the data set X, the goals, g, and the utility U are assumed identical. This definition describes what is done by a specific analysis. In order to assess how it is carried out, InfoQ is deconstructed into 8 dimensions. These are: (1) Data resolution, (2) Data structure, (3) Data integration, (4) Temporal relevance, (5) Generalizability, (6) Chronology of data and goal, (7) Construct and action operationalization and (8) Communication.

The goal, g, we are considering here is the use of customer surveys to determine drivers of customer satisfaction. The analysis we conduct, f, relies on Bayesian Networks as explained in the next section. The next section demonstrates our analysis using the ABC data set that is available from the web site of Kenett and Salini, 2012.

3. Bayesian Networks Applied to Customer Surveys

An early attempt to apply Bayesian Networks for the analysis of customer surveys was presented in Kenett and Salini (2009) and Salini and Kenett (2009). A survey with n questions produces responses that can be considered as random variables, $X_1, \ldots, X_n$. Some of these variables, q of them, are responses to questions on overall satisfaction,