Available online at www.sciencedirect.com
ScienceDirect
Information & Management 45 (2008) 164–174
INFORMATION & MANAGEMENT
www.elsevier.com/locate/im

Integrating the voice of customers through call center emails into a decision support system for churn prediction

Kristof Coussement, Dirk Van den Poel *

Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweekerkenstraat 2, 9000 Ghent, Belgium

Received 9 October 2006; received in revised form 7 August 2007; accepted 21 January 2008
Available online 17 March 2008

Abstract

We studied the problem of optimizing the performance of a DSS for churn prediction. In particular, we investigated the beneficial effect of adding the voice of customers through call center emails – i.e. textual information – to a churn-prediction system that only uses traditional marketing information. We found that adding unstructured, textual information into a conventional churn-prediction model resulted in a significant increase in predictive performance. From a managerial point of view, this integrated framework helps marketing-decision makers to better identify customers most prone to switch. Consequently, their customer retention campaigns can be targeted more effectively because the prediction method is better at detecting those customers who are likely to leave.
© 2008 Elsevier B.V. All rights reserved.

Keywords: Customer relationship management (CRM); Data mining; Churn prediction; Text mining; Call center email; Voice of customers (VOC); Binary classification modeling

1. Introduction

In the past, companies focused on selling products and services with little knowledge or strategy concerning the customers who bought the products. Today business is evolving from this 'product-centered' to a 'customer-centered' environment. Companies need to find ways to capture and enhance market share while reducing costs [7]. Consequently, existing companies must reconsider the business relationships with their customers [24].

Customer relationship management (CRM) is becoming a critical success factor in today's business environment [2,16]. Data mining is being implemented to gain customer knowledge from organizational data warehouses [35]. A way to manage customer churn is to predict which customers are most likely to leave and then target them with incentives to stay. Consequently, these IS support marketing-decision makers to generate marketing campaigns for the right customers. A field experiment by Burez and Van den Poel [9] has already shown that companies can boost profitability by shifting from mass to focused marketing strategies. It is more profitable to keep and satisfy existing customers than to attract new ones with a high attrition rate [26]. Identifying customers most prone to switch, is thus important [17]. In order to develop an effective customer retention program, the company must build a model that is as accurate as possible; indeed Van den Poel and Larivière [36] showed that a small change in retention rate can result in a significant change in profitability.

We decided it was necessary to incorporate the voice of customers (VOC) through call center emails into a traditional churn-prediction model in order to provide a better model: one with a higher predictive performance. The rapid development of IT and the Internet has made it easier for customers to communicate with the company. Call centers are expanding rapidly in scope, number and size [1], because many firms rely on them to address

* Corresponding author. Tel.: +32 9 264 89 80; fax: +32 9 264 42 79.
E-mail addresses: Kristof.Coussement@Ugent.be (K. Coussement), Dirk.VandenPoel@Ugent.be (D. Van den Poel).
URL: http://www.crm.UGent.be

0378-7206/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.im.2008.01.005

customer concerns and provide product information [25]. However, marketing managers tend to neglect this valuable information because (i) it is not directly applicable in a traditional marketing context, (ii) there is seldom in-house knowledge on how to convert this (textual) information into an analyzable form and (iii) no ready-to-use framework is available to integrate the information. We developed a DSS for churn prediction; it integrates free-formatted, textual information from customer emails with information derived from the marketing database. Although previous research used the VOC in understanding customers' needs and behavior (e.g. Refs. [10,11,21]), no prior work has used VOC in a churn-prediction model.

2. Methodology

Fig. 1 shows how the integration of information types in a churn-modeling system was achieved.

2.1. Data collection

Structured marketing information can be extracted from a common marketing database in which all transactional and marketing-related information has been stored. In contrast, call center emails are highly unstructured. Thus, extracting information from emails requires meticulous pre-processing to capture the relevant details for inclusion in a churn detection/prediction DSS.

2.2. Pre-processing

2.2.1. Data and text pre-processing

The structured information is internally available at a very low cost and available for pre-processing and integration into our model. However, the original emails are unformatted by nature. They are converted into a structured representation using the vector-space of Salton's SMART [31]: an email is represented as a vector of weighted frequencies of designated words. Thus emails are n-dimensional vectors, with n the number of distinct terms in the dictionary. Each vector component reflects the importance of the corresponding term with respect to the semantics of the email [6] and each component has a weight if the term is present or zero otherwise. Thus a collection of emails is represented as a term-by-email matrix. Fig. 2 shows the steps in this pre-processing phase whereby raw emails become a term-by-email matrix.

In the first step, *raw text cleaning*, special characters and punctuation are removed from words and spelling errors are corrected by comparing with words in a reference dictionary using a synonym data set. *Tokenization* converts the input stream into tokens or words. It uses blanks as delimiters for words which are then converted to lower case (*case conversion*). *Part-of-speech tagging* gives words their syntactic category: informative (nouns, verbs, adjectives and adverbs) or non-informative.

Next, terms are replaced by their stem, e.g. *connect* is the stem for *connected*, *connecting*, *connection*, *etc*., *Stemming* reduces the number of terms significantly [5] and increases

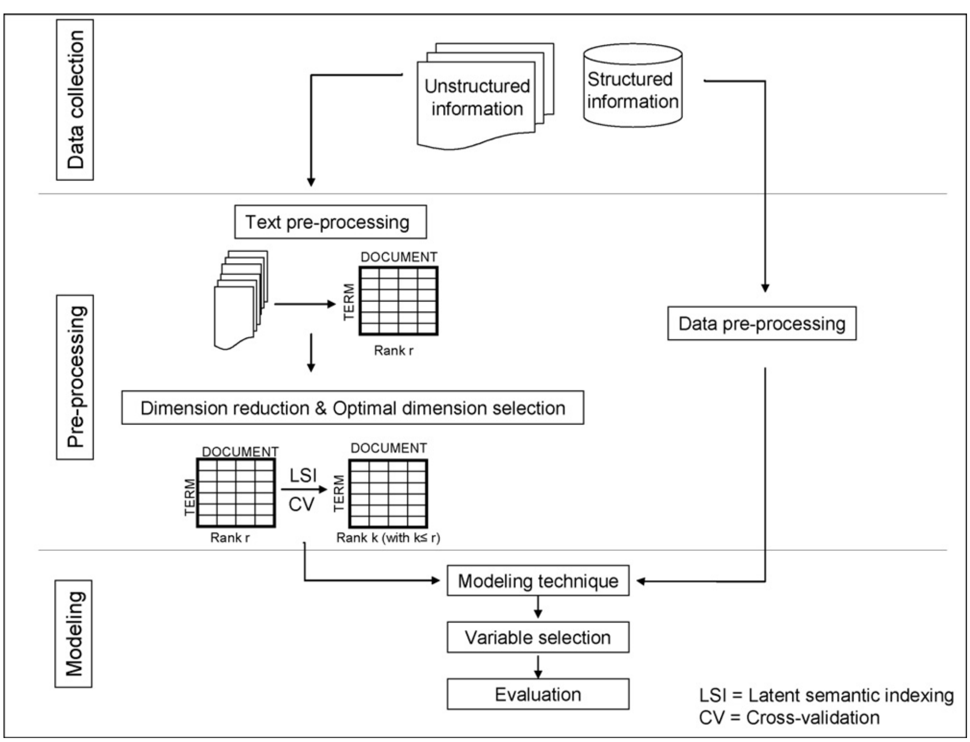

Fig. 1. An integrated churn-modeling system that uses structured, database-related information and free-formatted, textual information.

Download English Version:

https://daneshyari.com/en/article/554180

Download Persian Version:

https://daneshyari.com/article/554180

Daneshyari.com