Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 33 (2007) 181–191
Expert Systems with Applications
www.elsevier.com/locate/eswa

Extracting underlying meaningful features and canceling noise using independent component analysis for direct marketing

Hyunchul Ahn [a,*], Eunsup Choi [b], Ingoo Han [a]

[a] Graduate School of Management, Korea Advanced Institute of Science and Technology, 207-43 Cheongrangri-Dong, Dongdaemun-Gu, Seoul 130-722, Republic of Korea
[b] Department of Computer Science, Korea Military Academy, Gongnung-Dong, Nowon-Gu, Seoul 139-799, Republic of Korea

Abstract

As the Internet spreads widely, it has become easier for companies to obtain and utilize valuable information on their customers. Nevertheless, many of them have difficulty in using the information effectively because of the huge amount of data from their customers that must to be analyzed. In addition, the data usually contains much noise due to anonymity of the Internet. Consequently, extracting the underlying meanings and canceling the noise of the collected customer data are crucial for the companies to implement their strategies for customer relationship management. As a novel solution, we propose the use of independent component analysis (ICA). ICA is a multivariate statistical tool which extracts independent components or sources of information, given only observed data that are assumed to be linear mixtures of some unknown sources. Moreover, ICA is able to reduce the dimension of the observed data, especially noisy variables. To validate the usefulness of ICA, we applied it to a real-world one-to-one marketing case. In this study, we used ICA as a preprocessing tool, and made a prediction for potential buyers using artificial neural networks (ANNs). We also applied PCA as a comparative model for ICA. The experimental results showed that ICA-preprocessed ANN outperformed all the comparative classifiers without preprocessing as well as PCA-preprocessed ANN.

Keywords: Independent component analysis; Principal component analysis; Artificial neural networks; Direct marketing; Customer relationship management

1. Introduction

The Internet has changed the environment for management dramatically in many ways. Especially, the Internet has provided various ways for companies to build communication channels and relationships with their customers. Recently, new technologies such as WWW (World Wide Web), DW (Data Warehouse), and mobile communications have made it easier to collect a huge amount of data on customers and to provide personalized cues to attract them to make a purchase. Thus, in order to be a winner in severe competition with other competitors, the companies should deeply understand their customers and provide more sophisticated products or services by using up-to-date information technologies.

However, it is never easy to implement a system that facilitates a deep understanding of customers in a real-world business, although they are already equipped with state-of-art information technologies. There are two main reasons. The first reason is that there is too much information (i.e. information overload) on customers, so it is very hard to understand all the information in-depth. Moreover, most of the collected information is usually a record of simple observations although the useful information for the companies is more than just observations, but their underlying meaning. In particular, it is generally almost

* Corresponding author. Tel.: +82 2 958 3685; fax: +82 2 958 3604.
E-mail addresses: hcahn@kaist.ac.kr (H. Ahn), choies00@kma.ac.kr (E. Choi), ighan@kgsm.kaist.ac.kr (I. Han).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.04.016

impossible to interpret the observations to find the underlying meanings without the help of human experts.

The second reason is the fact that the data collected in an online environment usually contains much noise. When considering the principle of GIGO – Garbage In, Garbage Out – the noisy data set may lead the companies' strategies for their customers in a wrong direction. Consequently, extracting the underlying meaning from the observed data without prior knowledge (sometimes called 'blind source separation' in other engineering areas) and canceling noise from the collected data have been very important issues in data-driven marketing approaches.

This study proposes a novel technique for overcoming these obstacles – independent component analysis (ICA). ICA is a recently developed method which finds a linear representation of non-Gaussian data whose components are statistically independent (Hyvärinen & Oja, 2000). It is quite similar to principal component analysis (PCA), and PCA and its variations have been applied for similar purposes until now (Casarotto, Bianchi, Cerutti, & Chiarenza, 2004; Jutten & Herault, 1991; Karhunen, Pajunen, & Oja, 1998; Kim & Yum, 2005; Stetter et al., 2000; Zhu, Ye, & Zhang, 2005). However, ICA uses a higher-order statistical method while PCA uses a second-order method. Thus, more useful information may be extracted from the data in ICA than PCA (Back & Weigend, 1997; Cao, Chua, Chong, Lee, & Gu, 2003; Du, Hu, & Shyu, 2004; Fragos, Stergioulas, & Xydeas, 2003; Katsumata & Matsuyama, 2005; Kermit & Tomic, 2003).

As a result, ICA has been applied to a number of application cases including motion and image identification (Du et al., 2004; Katsumata & Matsuyama, 2005), demand estimation (Liao & Niebur, 2003), financial data analysis (Back & Weigend, 1997; Cao et al., 2003; Kiviluoto & Oja, 1998; Wu & Yu, 2005) and behavioral prediction (Fragos et al., 2003). However, there have been few studies that use ICA as a preprocessing tool of customer data for customer relationship management (CRM). In this study, we propose ICA as a method to extract underlying meaningful information from collected customer data without prior knowledge and also to remove useless noisy variables. To validate the usefulness of ICA, we applied it to a real-world one-to-one marketing case. In addition, we applied PCA to the same data as a comparative method to confirm the superiority of ICA compared to PCA. After these two techniques transformed the original data sets into the preprocessed ones, we used them as inputs for prediction models, artificial neural networks (ANNs). Finally, we compared the prediction accuracy of each method.

The article is organized as follows. In Section 2, we provide a brief explanation on the theoretical background of ICA and PCA. In the next section, ICA – the proposed method – will be explained in-depth. Section 4 presents the research design and experiments. In the fifth section, the experimental results are presented and discussed. The final section suggests the contributions and limitations of this study.

2. Theoretical background

This section introduces the theoretical background of ICA and PCA from the perspective of linear transformations. It will facilitate an understanding of the meaning of ICA and the difference between ICA and PCA.

2.1. Linear transformation

One of the fundamental problems in multivariate data analysis is finding suitable representations of measured data. Suitable representations mean transformations of the data that make a certain desirable feature of the data more accessible in the further analysis (Kermit & Tomic, 2003). When an n-dimensional data vector, denoted by $\mathbf{x}$, is considered, a general transformation of the data is represented as Eq. (1)

$$\mathbf{y} = f(\mathbf{x}) \tag{1}$$

where $\mathbf{x}$ is an n-dimensional vector $[x_1 \quad \cdots \quad x_n]^{\mathrm{T}}$, f is a function that performs the transformation and $\mathbf{y}$ is an m-dimensional vector $[y_1 \quad \cdots \quad y_m]^{\mathrm{T}}$ that represents transformed data with the desired feature illuminated.

Function f can represent various kinds of transformations. But, a linear transformation is most popularly used as a transformation function of the observed variables. In case of linear transformations, Eq. (1) can be transformed into Eq. (2)

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{2}$$

where $\mathbf{W}$ is the matrix that implies transformation.

As shown in Eq. (2), all computations are performed using a matrix in linear transformations, which means they are computationally easy to handle. This characteristic reduces the complexity of linear transformations as well as computational resources. Moreover, it also facilitates the interpretation of the results. Due to these advantages, linear transformations have been used for a long time in many application areas and various techniques of finding linear transformations have been developed. Principle component analysis (PCA), factor analysis, and projection pursuit are examples of well-known techniques for finding linear transformations.

In general, the techniques for linear transformations can be classified into two groups by the type of statistical information used for transformation (Hyvärinen, 1999b). One is second-order methods and the other is higher-order methods. The former methods use the variance within the data to estimate the transformation, while the latter methods use the information about the data density that is not contained in the covariance structure – cumulants and moments of order greater than two (Kermit & Tomic, 2003). Thus, to be brief, one might roughly summarize that higher-order methods are likely to provide more meaningful representations based on truly statistical independence while the second order methods represent the data