

Statistical analysis for customer profiling

Analisi del comportamento dei clienti di una azienda

Luisa Bisaglia¹

Dipartimento di Scienze Statistiche,
Università di Padova, e-mail: luisa.bisaglia@unipd.it

Margherita Gerolimetto²

Dipartimento di Statistica,
Università Ca' Foscari, e-mail: margherita.gerolimetto@unive.it

Bruno Scarpa¹

Dipartimento di Scienze Statistiche,
Università di Padova, e-mail: bruno.scarpa@unipd.it

Riassunto: La segmentazione della *Customer Base* e in particolare il *profiling* dei comportamenti dei singoli clienti rispetto all'utilizzo e fruizione dei prodotti e servizi caratterizza, ormai, la strategia di marketing di molte aziende. In particolare, le aziende che richiedono una continua interazione con i propri clienti (come banche, compagnie di assicurazione, società di telecomunicazioni, gestori di servizi internet ecc.) per prime hanno implementato sofisticati strumenti di classificazione della clientela per migliorare la relazione con i singoli clienti. Gli strumenti quantitativi maggiormente usati in questo contesto considerano le variabili disponibili principalmente in maniera statica, senza trarre vantaggio dalle caratteristiche dinamiche. Queste ultime, invece, sarebbero di aiuto per meglio descrivere il profilo dei diversi clienti, dal momento che questo è caratterizzato da realtà che si modificano spesso lentamente, ma a volte drasticamente nel tempo. Nel presente lavoro affronteremo questo tema proponendo una segmentazione della *Customer Base* di una società di telecomunicazioni basata sull'andamento del traffico dei singoli individui in un anno.

Keywords: profiling, segmentation, time series clustering

1. Customer profiling

The marketing departments of most modern companies offering products and services characterized by a continuous interaction with customers, such as banks, insurance companies, telecommunication or internet providers, spend huge efforts in classifying the customers in homogeneous clusters, in order to specify the relationship with each of them. Customer Segmentation is, in fact, a classical marketing strategy which has become much more effective since the progress of information technology allowed the analysts to use big amounts of data and fast algorithms of clustering. In this context, marketing strategies went also deeper in the personalization of the relationship with customers by analyzing the profile of each single customer and, automatically, offering to him what is believed to be the best for him in terms of contact channel, special offers, style in relation etc.

¹Address of correspondence: via Cesare Battisti, 241, Padova

²Address of correspondence: San Giobbe, Cannaregio 873, Venezia

Therefore, customer profiling has become one of the most challenging areas for statistical analysis and modelling in marketing research, particularly for companies, such as telecommunication (TLC) companies, that store in big data warehouse an enormous amount of information about the usage of product and services. Usually, all this information is, of course, aggregated to produce one (or more) dataset that is much more informative with respect to the specific goal of the marketing group for every programmed action.

Classical models for customer profiling (see for instance Berry and Linoff (2001)) are based on the analysis of all the available data, often using data mining tools (see for instance Azzalini and Scarpa (2004)), eventually creating new variables, such as index numbers measuring rates of decrease or increase of variables (e.g. number of phone calls) from one month to the other. However, the time dependency among single variables is not often considered as informative in profiling and segmenting customer base, although it can be very useful, especially if forecast is among the objectives. In other contexts, many statistical models have been proposed to get clusters using data from a longitudinal framework. Among others Piccolo (1990) and Maharaj (1996) first model time series and then define a distance in terms of the model parameters, Alonso *et al.* (2006) propose a clustering method based on the full probability density of the forecasts.

2. Telecommunication case

In this paper we analyze data from the customer base of a telecommunication company. Data are collected from 30.619 customers and we are interested in the segmentation of this population in homogeneous clusters, using the available variables. Some of them are “static”, such as gender, date of birth of the customer, activation date, tariff plan or value added service subscribed. Some of them are time varying characteristics, mainly related to the traffic behaviour observed for nine consecutive months, such as number, duration and value of outgoing calls, number and duration of incoming calls, number of sms, or number of call to the customer service. We cluster customers considering the entire trend of the traffic variables, by estimating some simple and easily interpretable models for each customer and using estimated parameter as element for a classical cluster analysis.

References

- Alonso A., Berrendero J., Hernández A. and Justel A. (2006) Time series clustering based on forecast densities, *Computational Statistics and Data Analysis*, 51, 762–776.
- Azzalini A. and Scarpa B. (2004) *Analisi dei dati e data mining*, Springer, Italia.
- Berry M. and Linoff G. (2001) *Mastering Data Mining: The Art and Science of Customer Relationship Management*, John Wiley & Sons, New York.
- Maharaj E. (1996) A significance test for classifying arma models, *Journal of Statistical Computation and Simulation*, 54, 305–331.
- Piccolo D. (1990) A distance measure for classifying arima models, *Journal of Time Series Analysis*, 25, 449–465.