EDITORIAL

Special issue on data analysis and classification in marketing—preface by the guest editors

Daniel Baier · Reinhold Decker

For more than five decades, marketing has been an important field of application for a multitude of data analysis and classification methods. The sophisticated collection of adequate data from and about customers and competitors as well as the purposeful application of methods like cluster, discriminant, factor and regression analysis, multi-dimensional scaling and association analysis have become widely accepted standards in the treatment of marketing problems. Decision support is provided in that way for market segmentation, market structuring, target marketing/advertising, new product development, pricing, positioning, and cross-media communication, to mention only a few. The lessons learned from small and large-scale applications have motivated marketing researchers to develop dedicated methods and to further improve existing methods with regard to their fields of application. Since the theoretical and method-ological standards in the fields of interest are very high today, promising adoptions of data analysis and classification methods from other application fields or directly from theory are challenges of continuous importance.

The present Special Issue of ADAC highlights recent developments regarding the mentioned aspects in marketing dealing with topics like

– Innovative methods for data analysis and classification in marketing and marketing research,

D. Baier (✉)
Chair of Marketing and Innovation Management, Brandenburg University of Technology Cottbus,
Erich-Weinert-Strasse 1, 03046 Cottbus, Germany
e-mail: daniel.baier@tu-cottbus.de

R. Decker
Department of Business Administration and Economics, Bielefeld University,
P.O. Box 10 01 31, 33501 Bielefeld, Germany
e-mail: rdecker@wiwi.uni-bielefeld.de

springer

– In-depth investigations of established data analysis and classification methods in marketing and marketing research, as well as,
– The development and empirical verification of analysis and classification methods for new types of marketing data.

All submissions that fitted to the thematic guidelines published in the corresponding call for papers passed a peer-review process. The seven finally accepted papers illustrate the broad spectrum of modern data analysis and classification in marketing. They vividly reveal that current research in this area increasingly leaves the beaten tracks.

The paper by *Daniel Baier*, *Ines Daniel*, *Sarah Frost*, and *Robert Naundorf* discusses how the ever increasing amount of uploaded photographs by consumers in social networks or during online interviews can be used for marketing purposes. With a focus on lifestyle segmentation, they show how the content of uploaded photographs can be used to characterize consumers and consequently group them into homogenous groups. The paper starts with an overview of available algorithms to represent digital images by extracted features and how to measure image (dis)similarities. Then, these algorithms are applied and compared to traditional analyses in this field. A new software package for image data analysis and classification (IMADAC) developed by the authors is presented and applied for this purpose.

Paola Cerchiello and *Paolo Giudici* propose a new approach for opinion spam detection in social networks. They discuss why support vector machines and naive Bayes classifiers have difficulties with this problem and show in an empirical application how the new approach works. The main idea is to decide whether a new opinion (i.e. a newly posted document) stems from a new consumer (a person not having written an opinion so far) or from an already known consumer (assuming that this is a potential spammer who postes more than one opinion to influence the others). Their new approach—a combination of classication trees, the non-parametric Kruskal–Wallis and Brunner–Dette–Munk tests—shows promising results.

The contribution by *Wolfgang Gaul* and *Dominic Gastes* discusses consistency improvement techniques for the analytic hierarchy process (AHP), a popular multicriteria decision-making approach, and shows that not all of them are helpful for computing acceptable weights for the determination of the underlying overall objective function. The paper starts with a short description of the AHP approach and the available consistency improvement techniques. Then, a simulation study shows to which extent the adjusted matrices are still different from the underlying consistent 'true' matrices. Additionally, a possibility is described for supporting a judging person to report a consistent paired comparison matrix.

Maria Iannario, *Marica Manisera*, *Domenico Piccolo*, and *Paola Zuccolotto* discuss how useful information for marketing management can be obtained by combining the results from so-called CUB models (i.e. convex combinations of discrete uniform and shifted binomial distributions) and algorithmic data mining techniques, both together with variable importance measurements from random forests methodology. A case study on sensory evaluation of different varieties of Italian espresso is presented to illustrate the theoretical considerations. Among others, the authors empirically show that the opinion regarding the suitability of a person for drinking a certain coffee is

very important in determining the overall sensory satisfaction and that those who are satisfied with a coffee tend to consider this coffee suitable for sophisticated/luxurious persons.

A new approach to customer satisfaction evaluation based on three-way factor analysis is described by *Caterina Liberati* and *Paolo Mariani*. They additionally introduce a reassessment technique to adjust the three-way solution according to the representative quality of particular points. A special feature of their approach is its extension to the multi-period case which allows a dynamic view on the evolution of customer satisfaction. This is achieved by the identification of customer trajectories and the derivation of synthetic measures of customer (satisfaction) evolution. The effectiveness and usefulness of their approach is demonstrated using a large-scale data set on customer satisfaction in the banking sector. Among others, the authors explore main aspects of customer satisfaction and identify areas for possible improvements.

Vera L. Miguéis, *Dirk Van den Poel*, *Ana S. Camanho*, and *João Falcão e Cunha* introduce a new approach to churn prediction which measures the similarity of the product sequence first purchased with churner and non-churner sequences. These sequences of first purchase events are modelled as a Markov process. Unlike most of the previous churn models proposed in the literature, the authors propose a prediction model which identifies those customers who probably are going to (partially) leave the retailer. In their empirical study they compare the predictive performance of logistic regression and random forests based on a large sample of customers provided by a European retailing company. The presented results highlight the relevance of the proposed model, since the performance of the models including the sequence likelihood variables is higher than the performance of those not including these variables. Furthermore, the results suggest that the logistic regression technique may outperform the random forests technique in similar settings.

The paper by *Takanobu Nakahara* and *Katsutoshi Yada* focuses on the systematic analysis of shopping path data resulting from RFID-based customer tracking in a Japanese supermarket. By using information on the sequence of visiting each product zone they investigate how the visiting of and staying in a particular area affect individual purchasing behaviour. To discover useful knowledge for the store management the shopping path data is transformed into sequence data including information on visit sequences and staying times. Frequent sequence patterns are extracted using the so-called linear time closed itemset miner (LCM) algorithm. Then, by using these patterns as descriptive variables for discriminant analysis the authors identify contributing factors characterizing each set of customers considered.

The submissions and the finally accepted papers show that the methodological standards in the present field of interest are very high today. The editors wish to thank the authors and the reviewers for their valuable work.