The Use of Neural Networks in Marketing: Market Segmentation with Self Organising Feature Maps

Sara Dolnicar
Institute for Tourism and Leisure Studies, Vienna University of Economics and Business Administration
Augasse 2-6, 1090 Vienna, Austria
Telephone: 313 36 /48 50, Fax: 317 12 05
e-mail: sara.dolnicar@wu-wien.ac.at

Abstract

A taxonomy of tourists visiting Austria during the summer season is constructed using Kohonen's self organising feature maps and conventional k-means cluster analysis. 8000 tourists visiting Austria in the summer season of 1994 indicated the importance they attach to 22 vacation aspects that are used as active clustering variables. The resulting market segments from both methodological approaches are described. Comparisons are made between self organising feature maps and k-means clustering concerning the final results, and the variability of final results. Self organising feature maps turn out to be an interesting alternative to conventional k-means clustering and furthermore bear the potential of unveiling additional precious information for market segmentation studies.

Introduction

Tourism demand is increasing. Nevertheless numerous travel destinations have to cope with a decreasing number of visitors. Market segmentation is one of the strategic marketing instruments that enables companies in the tourism industry as well as regional tourism organisations to get some insight into the market they serve, thus offering the possibility to analyse customers, and adapt marketing actions accordingly. Market segments are groups of customers that are homogenous in some predefined variables within the segment and differ strongly from each other over the groups. Segments once identified form the basis for well-targeted successful marketing actions. Segment analysis itself is based on data obtained by questioning customers or potential customers as well as on the use of methodology that enables data analysis with large amounts of complex market data. Calculations are usually made applying factor analysis, discriminance analysis and cluster analysis. The latest development in this field of research shows an increasing number of successful attempts to use neural network technology for solving data analysis problems. Caudill [1993] and Kohonen [1990] list numerous reasons that make the use of self organising feature maps (SOFMs) interesting for the evaluation of empirical survey data: rapid, continuous learning, fast adaptation to changes in the input data structure, network training without need for correct output results (unsupervised learning), suitability for large amounts of data, possibility of transfer of identified classes to new cases.

Numerous variables have been used since market segmentation became popular in research and business. Generally, a trend towards multivariate approaches is observable, including demographic, socio-economic, behavioural, life-style and psychographic variables. These complex approaches lead to more informative and useful results, a fact that has already been proven by Abbey [1979] comparing demographic and life-style variables in segment identification. Research interest not only shifted from fairly simple to complex approaches but also from more to less superficial information. The importance of psychographic variables is therefore increasing. Numerous surveys have successfully searched for underlying relevant dimensions of personality [Plog 1974, Dalen 1989, Mayo 1975, Shih 1986, Calantone & Johar 1984). The tendency towards underlying aspects also lead to a change of the function of demographic and socio-economic variables that are now primarily used for describing identified clusters as comprehensively as possible instead of extracting group membership on the basis of such variables.

This paper centres around the use of SOFMs for the identification of market segments. The result is a psychographic a-posteriori-segmentation of tourists visiting Austria. Two methods are used and compared: k-means clustering and Kohonen's SOFMs. This empirical, artificial and polythetic classification [Bailey 1994] results from analysis of answers given by 7864 tourists spending their vacation in Austria at the time of the interview, to the question which aspects of a vacation are important to them.

Data Set

The Austrian Guest Survey is conducted every three years, questioning tourists visiting Austria in the summer and in the winter season. Respondents are asked to answer questions concerning demographic and socio-economic status, behaviour during the stay, and psychographic aspects. The data set underlying this segmentation study is taken from the Austrian Guest Survey in the summer season of 1994. Only the section containing psychographic information was used as a basis for segment identification: the respondents stated the importance (high or low) they attach to 22 aspects of vacation, as e.g. security, rest and relaxation, comfort, sports. The binary data is a matrix with 7864 cases (respondents) and 22 variables (importance attached to the aspects listed).

Methodology

As mentioned above, two methodological approaches have been used: k-means clustering [Everitt 1993] and SOFMs [Kohonen 1984]. Solutions resulting from k-means clustering (nine clusters) provide a comparison for SOFM results. The starting point for the SOFM training run with nine nodes (arranged in a three times three matrix) was a random solution of the prototype vectors with 100 trials. Learning rate was reduced from 0.05 to 0.00001 in steps of 1/100sth, the largest neighbourhood size was two with an learning interaction for neighbouring prototypes of 0.5. At each level of the learning rate the net was allowed to learn for 10 epochs. Average values of results were used for comparison among the methodologies.

The values of heterogeneity, simplicity and violations were recorded for every 10 epochs during the training run of the net. Heterogeneity meaning the sum of squared euclidean distances of each data vector to the best representing prototype, divided by the total number of data points. Thus indicating better quality of the classification solution with lower values. The simplicity of the SOFM is calculated by the sum of squared euclidean inter-prototype distances between adjacent prototypes. The idea of using violations to extract information on the quality of the SOFM solution is that adjacent prototypes in the final resulting matrix should have lower euclidean distances to each other than neighbours of higher order. To obtain the violations value for a SOFM solution, pairwise prototype distances are computed and compared to the neighbourhood relation. The higher the violations in percent, the higher is the number of prototypes pairs that violated the rule of correct ordinal distance within the matrix arrangement.

Results

Psychographic Taxonomy of Tourists visiting Austria

Group memberships indicated for each case processed in k-means clustering, and the recall run of the SOFM, formed the basis for the identification of travel segments. The interpretation of solutions using both methodological approaches indicated the existence of 10 tourist segments, three of which turned out to be subgroups of one more general group. The following segments resulted from analysis of active variables:

'Modests' attach very low importance to each one of the vacation aspects stated, contrarily to 'I-expect-it-alls' with very high expectations.

'Sportsfreaks' are defined through their extremely high need for physical activity during their stay. Sports are very important for more than 90 % of travellers in this group. Besides sports, this segment emphasises adventure, nature, protection of the environment and the change of pace during the vacation more than the average does. Very few members of this group search for relaxation, comfort and perfect organisation. To sum it up, 'sportsfreaks' describe themselves as very active and adventurous and need a destination that offers the necessary environmental conditions.

Culture, local population and nature are very important for more than 80 % of the 'culture Seekers'. This segment has a primary motive for visiting Austria: getting to know the culture. As far as importance statements are concerned this is the one and only activity during the vacation, culture meaning more than sightseeing alone. 'Culture Seekers' also want to learn about the destination county and the people living there.

The profile of 'pleasure seekers' is dominated by vacation features such as entertainment, fun, culture and excitement. As long as they enjoy themselves during their stay, the aspects such as nature, environmental protection and local population are of low interest to them. This fact, and the higher level of socialising distinguishes them from 'culture seekers'. 'Pleasure seekers' love to have fun, and they love it even more, when they are in pleasant company and stay in a comfortable place.

The group with less then 10 % members stating that sports, health and water and sun are important has been labelled 'active philanthropes'. Highly ranked items include culture, excitement, fun, entertainment, and romantic atmosphere. A segment with high expectations concerning the travel destination, that shows enthusiasm for activity and social contact.

Less active but more social: 'socialisers'. Almost all members of this group state that fun is important. Relaxation, enjoyable company, cosy and familiar atmosphere and feeling free and easy are the basis for a vacation to remember for more than 80 % of the respondents in this group. It seems that swimming and sunbathing are the main activities in this group.

Finally, the largest segment of tourists in Austria can be named 'rest seekers'. Therefore both methods identified subgroups within this group that show specific interest although falling into the same general profile of travel expectations: rest and relaxation, nature, environmental protection and change of place are crucial to them. Members of the subgroup 'radical rest seekers' have no need for comfort, while 'comfort and rest seekers' can not imagine relaxing without comfort. 'Romantic rest seekers' are the only group, with a high proportion of group members caring about romantic experiences during their stay.

The distribution of respondents among these segments is shown in the following table.

Table 1: Number of respondents per identified segment resulting from the SOFM recall run and k-means clustering

	SOFM	**k-means**
Sportsfreaks	747	775
Modests	843	795
I-expect-it-alls	820	759
Socialisers	844	855
Active Philanthropes	838	823
Culture Seekers	914	928
Pleasure Seekers		687
Radical Rest Seekers	897	1067
Comfort & Rest Seekers	921	1175
Romantic Rest Seekers	1040	

It can be observed that most of the tourist segments have been identified by both methodological approaches. 'Pleasure seekers' and 'romantic rest seekers' each resulted from only one method; a result that purely methodological quality criteria are not able to provide. Heterogeneity of both solutions was nearly identical, only the interpretation of the segment profiles demonstrates the difference.

Comparison of SOFMs and k-means clustering

In a first step, final values of heterogeneity, simplicity and violations are compared. In fact, a number of other comparisons have been performed, however, comparisons of final results and reliability were found to be most relevant for the sake of the present presentation: As far as final results are concerned, k-means clustering results in slightly higher average heterogeneity. Specified in percent, heterogeneity values resulting from SOFM calculations beat k-means by no more than 0,28 %. Differences in final classification heterogeneity are therefore minimal and do not allow the conclusion of SOFM superiority in data clustering.

However, simplicity and violations are not necessary the conditions for arriving at a valid tourist segmentation. Nevertheless comparison in terms of both criteria is of high interest, since SOFM enable two-dimensional mappings of resulting clusters implying correct neighbourhood relations. Knowing which segment is more or less similar to other groups can be very helpful for interpretation of cluster profiles and drawing of managerial consequences. The feature of correct topographical mapping would therefore lead to an improvement in segmentation analysis as compared to cluster analysis. There is no doubt that SOFMs will lead to better results for the simplicity and violations values, since conventional clustering does not update neighbouring centroids during the classification process. But how much better are the SOFM results ? In this study, where the average values of ten SOFM and 20 k-means clustering calculations were compared, SOFM surpassed k-means clustering by 8.6 % in terms of simplicity and 2.6 % in terms of violations.

To sum up the comparison of final results: although SOFM render lower simplicity and violations values, thus mapping tourist segments more correctly on a two dimensional space with ordinal distances due to neighbourhood updating during the training run, classification quality as indicated by heterogeneity does not differ by more than one percent. It therefore seems worth investing further work in classification analysis with SOFM, since it seems to offer some advantages without substantial sacrifice in terms of the primary goal: classification quality.

Comparing the best results of each applied method may not provide enough information for the decision, which approach to use in the future. Another crucial point is variability or reliability of final results. Certainly no one would prefer a method that has the potential of achieving extremely favourable levels of heterogeneity, simplicity and violations once in a hundred trials. We therefore aim at using a method that computes a classification, giving us the security that the next hundred calculations would not render substantially different results. So, how reliable are the results compared before ?

To answer this question, the best and worst results were compared for each method. As far as the lowest (best) heterogeneity values are concerned, both k-means clustering and SOFM render equal results (the k-means result is higher by 0.004%). Differences in the high level heterogeneity results diverge slightly more (the k-means result is higher by 0,7 %). The average deviation of the final results from the mean value of the k-means clustering calculations in percent (0.19 %) surpasses deviation of SOFM (0.06 %) by 0.13 %. As far as reliability of classification quality is concerned, both methods are equally good in providing results of low variability

As far as simplicity and violations are concerned, we again expect SOFMs to lead to superior results. Maximum (58.7 for k-means and 51.2 for SOFM) and minimum simplicity (46.7 for k-means and 43.3 for SOFM) values, and the span between both (12 % for k-means as compared to 8% for SOFM) are lower when calculated with SOFMs. The best violations result achieved by k-means clustering is 18.8 % as compared to 10 % resulting from SOFM experiments. The difference between the respective maximum values is 3.8 % in favour of SOFMs, always bearing in mind that the k-means clustering algorithm does not aim at optimising simplicity and violations. In this case k-means clustering results vary less that SOFM results do, but on a overall higher level. This fact is also demonstrated by comparing the deviation from the mean value: k-means surpasses SOFMs by 5.8 %. It must be mentioned at this point that violations values in general vary strongly: when the last training epoch of the SOFM results in a violations percentage of 50, the interval of confidence reaches from 30 % to 70 %. In case of k-means lower and upper borders would be 35 % and 65 %, respectively. Fluctuation of violation values in general is high. Nevertheless, k-means clustering is more reliable on a higher value level.

Conclusions

Extraction of tourist types of guests visiting Austria during the summer of 1994 indicates the existence of eight basic groups with typical travel expectation profiles: ‘modests’, ‘culture seekers’, ‘pleasure seekers’, ‘active philanthropes’, ‘socialisers’, ‘sportsfreaks’, ‘I-expect-it-alls’ and ‘rest seekers’. ‘Rest seekers’ may be split up into ‘radical rest seekers’, ‘comfort & rest seekers’ and ‘romantic rest seekers’ who are discriminated by extreme interest in a few single variables.

Depending on training run parameters, SOFMs have the potential of, on the one hand, rendering equally good classification results as the k-means clustering algorithm, and mapping the neighbourhood relations of tourist types in two-dimensional space correctly on the other hand. Nevertheless, there is a trade-off relation between these two goals. Continuing the training run of a SOFM with a neighbourhood of zero will make SOFM a competitive classification tool with less optimal neighbourhood preservation in the final mapping. Ending the training run at a neighbourhood of one will lead to very good topographical mapping (the ten nearest neighbours were mapped as neighbours correctly in an experiment with a 4*5 prototype SOFM) and substantially higher heterogeneity levels than those achieved by k-means clustering.

The SOFMs do represent an interesting alternative to conventional clustering methodology. The ability to cope with the main data analysis goal, to identify clusters of similar cases, is comparable to k-means clustering. However, it appears that at this stage of studies, the application of SOFMs for classification analysis is not sufficient user-friendly enough at this point, as different data set features require a number of experiments in order to adapt training run parameters to the data set characteristics. It can be expected that more detailed research will diminish the trade-off relation between optimal classification and optimal mapping. A numeric quality indicator should be found combining heterogeneity, simplicity and violations in order to simplify the choice of the optimal solution before interpreting expectation profiles qualitatively.

References

Abbey, James R. ‘Does Life-Style Profiling Work ?’, *Journal of Travel Research, 18 (1), Summer 1979*, p 8-14.

Bailey, Kenneth D. *Typologies and Taxonomies: An Introduction to Classification Techniques.* Sage University Paper series on Quantitative Applications in the Social Sciences. Thousand Oaks: Sage, 1994.

Caudill, Maureen. *Neural Networks Primer*. 3rd ed. San Francisco: Miller Freeman, 1993.

Calantone, Roger J. & Johar, Jotindar S. ‘Seasonal Segmentation of the Tourism Market Using a Benefit Segmentation Framework’, *Journal of Travel Research. 23 (2), Fall 1984*, S 14-24.

Dalen, Eric. ‘Research into values and consumer trends in Norway’, *Tourism Management. September 1989*, S 183-186.

Everitt, Brian S. *Cluster Analysis*. New York: Halsted Press, 1993.

Hruschka, Harald & Natter, Martin. *Using Neural Networks for Clustering-Based Market Segmentation*. Forschungsbericht / Research Memorandum No. 307, Vienna University of Economics and Business Administration, September 1992.

Kohonen, Teuvo. ‘The Self-Organizing Map’, *Proceedings of the IEEE, Volume 78, Number 9, September 1990*, S 1464-1480.

Kohonen, Teuvo. *Self Organization and Associative Memory*. New York: Springer-Verlag, 1984.

Maren, Aliana J., Harston, Craig T. & Pap, Robert M. *Handbook of Neural Computing Applications.* San Diego: Academic Press, 1990.

Mayo, Edward. ‘Tourism and the National Parks: A Psychographic and Attitudinal Study’, *Journal of Travel Research. 14 (1), 1975*, 14-17.

Mazanec, Josef A. ‘How to Detect Travel Market Segments: A Clustering Approach’, *Journal of Travel Research. 23 (1), Summer 1984*, S 17-21.

Mazanec, Josef A. & Zins, Andreas. ‘Eurostyles and Socio-Targets as Guest Segments: Selected Findings’, *Revue de Tourisme - The Tourist Review - Zeitschrift für Fremdenverkehr. 2, 1992*, S 5 - 8.

Mazanec, Markus. SOMnia 1.4.1, Revision A. 1995.

Plog, Stanley C. ‘Why Destination Areas Rise and Fall in Popularity’, *The Cornell H.R.A. Quarterly. 14 (4), 1974*, S 55-60.

Plog, Stanley C. ‘Understanding Psychographics in Tourism Research’ in: Ritchie, J.R. Brent & Goeldner, Charles R. (Hrsg.). *Travel, Tourism, and Hospitality Research*. New York: John Wiley & Sons, 1987, 203-213.

Plog, Stanley C. *Leisure Travel - Making It a Growth Market ... Again!* New York: John Wiley & Sons, Inc., 1991.

Schewe, Charles D. & Calantone, Roger J. ‘Psychographic Segmentation of Tourists’, *Journal of Travel Research. 16 (3), Winter 1978*, S 14-20.

Shih, David. ‘VALS as a Tool of Tourism Market Research: The Pennsylvania Experience’ , *Journal of Travel Research. 24 (4), Spring 1986*, S 2-11.