3

3.2
Cases

3.2.1 SEGMENTATION, CLUSTERING

INTRODUCTION

Segmentation is often defined as 'being able to discern a group of entities with common characteristics'. From a mass of data one or more useful groups are identified. Examples might be customer groups or demographic regions. Segmentation is closely related to classification and is very useful for marketing purposes. More on the technical details of segmentation and classification can be found in Section 6.2.6, Clustering.

Advertising strategy discovery: publisher

Peter van der Putten[1]

In this section we describe a case in which a large Dutch publisher, De Telegraaf, uses customer profiling and segmentation to promote advertisement sales. In a marketing context, it is often tacitly assumed that customer databases are mined for knowledge. However, due to their rich nature (many fields) surveys are even more likely candidate data sources to be mined. The Telegraaf case is a very good example of this approach.

Business problem: promoting advertisement sales

De Telegraaf is one of the major publishers in the Netherlands. The Marketing Services Department is responsible for marketing of the largest Dutch daily newspaper 'De Telegraaf' and the magazines of the 'Telegraaf Tijdschriften Groep', which include various titles on music, lifestyle, sports, cars, etc. that are aimed at different target groups (like Autovisie, Elegance, Privé, Man, Hitkrant, Oor). A major problem for a publisher is that they only get contacted at the end of the advertisement chain, when the decision on where to advertise has already been made. Typically, for a certain campaign an advertising agency will design a media strategy and an appropriate communication message and hire a media agency to work out a detailed media plan and buy advertising space. By supplying large potential advertisers with high quality and serendipitous information about their target groups, De Telegraaf can get involved and grab some attention in the very beginning of the advertisement chain, maybe even before an advertising agency has been contacted. Furthermore, these target group reports are a win-win loyalty marketing tool to strengthen the relationship with major advertisers.

The target group reports are based on the so-called SUMMO Dutch National Media Survey, a rich collection of data on lifestyle, product and media consumption of Dutch consumers, which is carried out on an annual basis. In all major countries around the world similar national media surveys are available. For the SUMMO survey over 13,000 consumers are interviewed. The surveys contain over 5,000 answer codes (possible answers). Since 1994 De Telegraaf uses data mining technology to quickly discover relevant, interesting and surprising knowledge about target groups (see also [Putten, 1999]).

The standard statistical methods that De Telegraaf used before (cross tables) only offered the opportunity to calculate the frequencies of answers for a certain specific question, for example the frequency of Nike customers versus age or versus choice of newspaper. Given the enormous number of questions and

....

1   Drs P. van der Putten,
pvdputten@hotmail.com,
pvdputten@liacs.nl, Leiden Institute
of Advanced Computer Science,
Leiden University, Leiden, The
Netherlands

answers in the survey, this approach leads to practical problems. There is no time to screen all relevant questions and answers in the survey. Furthermore, the marketeers of De Telegraaf lack a priori knowledge about the target group. So even if unlimited time was available, even the most experienced marketeer might miss some important profiling characteristics of a target group. In practice, marketeers only have very limited time to prepare and present a report (one to two days!) and the goal is to surprise the potential advertiser with new knowledge, which is quite ambitious, if the client has years of experience with the target group. So it was desirable that instead of asking a marketeer to generate and test all interesting hypotheses, a mining tool should be used to generate potentially interesting profiling characteristics and segments [2].

DATA MINING SOLUTION: PROFILING AND SEGMENTATION

Data mining profiling analyses offer a solution to the problem. First a target group needs to be selected by use of standard strict criteria or fuzzy matching criteria, e.g. 'Sneaker=Nike' or 'Strict: Female, Fuzzy: around 50, highly educated, likes gardening & cooking, etc.'. Then the profiling engine uses standard statistical tests to generate a profile of deviating characteristics, by computing all frequency distributions of answers in the target group and comparing these frequencies with a reference group, for example the whole Dutch population. As an example the top profiling characteristics of vodka drinkers are displayed in Figure 1.

Figure 1

Target group profile for vodka drinkers. For example, watching science fiction movies is one of the most profiling characteristics, because 66,8% of vodka drinkers likes these movies, compared to 38,4% of the Dutch population, resulting in a selectivity index of 28,4 (numbers slightly changed for copyright reasons).

The target group profile offers a description of the average vodka drinker. However, this is only part of the solution. It may well be that the average vodka drinker does not exist, but that there are at least a small number of different vodka drinker types. The borders between these segments are typically not strict. Each of these customer types needs to be addressed with a different communication strategy (in other words: there are always opportunities for De

Telegraaf). Such a differentiated strategy results in both a higher efficiency and effectiveness. By using interactive data mining segmentation tools, the marketeer can discover which segments exist within the target group.

The segmentation process is carried out in a number of steps. First the target group, say all vodka drinkers, are spread out randomly on a flat plane. During the step-by-step segmentation process each vodka drinker moves a small distance towards those neighbors in the plane whose characteristics match best.

The end result is a plane with several segments (Figure 2): within a segment customers are as similar as possible and the segments are as different as possible. The match is not determined on just two variables such as gender and age, but on all questions and answers from the survey, or any subset of these questions, if desired.

The profiles of the various segments tell a different story from the profile of the average vodka drinker. Three segments can clearly be distinguished.

– A group of teens and adolescents, both boys and girls, who drink vodka with orange juice and cola, like to go out, watch videos and are interested in house music. This group does not read too much, apart from De Telegraaf and teen music magazines, but they watch advertisements in the cinema and in the street and they like to watch TV.
– A group of twenty-something year olds, students and young couples, who live in rented apartments and buy products like baby shampoo, fresh vegetables and cat food. They read more liberal newspapers like the Volkskrant and opinion magazines.
– A group of older business men; they own credit cards and other financial products, they drive around in lease cars more often and read conservative newspapers and magazines. They don't watch TV that much.

It is obvious that the profile of the average vodka drinker only presents a limited view. The segmentation exercise offers a richer view of the target group, which can contribute to better brand positioning and a distributed media strategy.

Vision for the future

For the past seven years De Telegraaf has been using descriptive data mining, and more specifically profiling and segmentation, as a major instrument for offering services to potential and actual advertisers, resulting in better advertisement sales. The success of this formula has been the integration of data mining in a core business process. De Telegraaf has a clear vision of how the data mining results should be exploited. The impact of the technical implementation has been minimized by updating the central survey database on a yearly basis.

Of course this customer profiling and segmentation approach can be applied (and has been applied) for a multitude of other purposes, not limited to surveys, such as insurance risk analysis, policy research, market basket analysis, customer relationship management, crime analysis and medical discovery.

From an application point of view it might be interesting to spend some future technical research on extending the rather straightforward univariate profiles to multivariate profiles that are generated by association rules and decision trees or non prepositional profiles that can be constructed by inductive logic programming algorithms. However, although these profiles are more expressive and complex, users might still prefer simple profiles! For segmentation it might be interesting to do some technical research on more immersive visualization technology such as three-dimensional segmentation and visualization. Again, the simple two-dimensional solutions will be difficult to defeat in terms of clarity.

Finally, for descriptive data mining tasks such as profiling and segmentation, the richer the data, the better it is. This is in contrast to prediction tasks that suffer more from the curse of dimensionality. However, rich data is expensive, and, generally, rich data is known only for subsets of customers. A new hot field in data mining concerns data fusion methods that allow us to enrich entire customer databases with survey information that is only available for a sample, in other words, carrying out a virtual survey with each customer [Putten, 2000]. If this technology becomes mature, a whole new arena for segmentation will evolve.

References

– Putten, P. van der. (1999). Datamining in Bedrijf. Informatie en Informatiebeleid **17**:3
– Putten, P. van der. (2000). Data Fusion: A Way to Provide More Data to Mine. Proceedings 12th Belgian-Dutch Artificial Intelligence. Conference BNAIC'2000, November. De Efteling, Kaatsheuvel, The Netherlands