

Product-Driven Data Mining

Wei Lu

Manifold Data Mining Inc.

The behaviour of consumers is believed to be influenced by many factors. Some of these factors include the individuals culture, social status, lifestyle and attitudes. Understanding how these complicated and interrelated factors drive the consumer is the primary goal of Manifold Data Mining. The question posed to the group was to 1) find an algorithm that predicts the likelihood of consumers to respond favorably to a given product. In addition, once this prediction is made for a given consumer the group was also asked to 2) develop a second algorithm that infers other statistical information regarding the consumer.

Manifold Data Mining has developed innovative demographic and household spending pattern databases for six-digit postal codes in Canada. Their collection of information consists of both demographic and expenditure variables which are expressed through thousands of individually tracked factors. This large collection of information about consumer behaviour is typically referred to as a *mine*. Although very large in practice, for the purposes of this report, the data mine consisted of m individuals and n factors where $m \simeq 2000$ and $n \simeq 50$. Ideally, the first algorithm would identify a few factors in the data mine which would differentiate customers in terms of a particular product preference. Then the second algorithm would build on this information by looking for patterns in the data mine which would identify related areas of consumer spending.

To test the algorithms two case studies were undertaken. The first study involved differentiating BMW and Honda car owners. The algorithms developed were reasonably successful at both finding questions that differentiate these two populations and identifying common characteristics amongst the groups of respondents. For the second case study it was hoped that the same algorithms could differentiate between consumers of two brands of beer. In this case the first algorithm was not as successful as differentiating between all groups; it showed some distinctions between beer drinkers and non-beer drinkers, but not as clearly defined as in the first case study. The second algorithm was then used successfully to further identify spending patterns once this distinction was made. In this second case study a deeper factor analysis could be used to identify a combination of factors which could be used in the first algorithm.