



## Data Mining Provides Retail Understanding

Analyzing customer data from all the possible sources - including the Web - can provide retailers with the intelligence to profitably influence their customers' behavior.

*Integrated Solutions for Retailers, February 2001*

by Stephanie Roussel-Dupre

Retailers collect terabytes upon terabytes of information every day - anything from transactional data, to demographics, to product sales based on seasons. But what do they do with it all once it is neatly organized into a database? The concept of data mining is just as it sounds. Companies drill holes through 0s and 1s to come up with relationships and patterns in customer habits. To a retailer this information can be more valuable than mining for gold, because the results are almost a guarantee. The data mining process used to be a highly technical process requiring mathematicians to build the analysis for companies. But today's data mining technology offers retailers the tools they need to make sense of their customer data and apply it to business. Mark Smith, president of Quadstone, a predictive marketing software company, and Peter Urban, senior research analyst at AMR Research, discuss the advantages of analyzing data from all sources and customer channels - including the Web.

### *What are the best sources of customer information for retailers?*

The best source is turning POS (point of sale) transaction data into measurements of customer behavior. The problem for many retailers is that they lack any information on specific customers, and hence are trapped analyzing data at the product and basket level. The rise of loyalty programs, mail order, and the Internet has provided retailers with real access to customers for the first time. This allows retailers to study the purchase behavior of customers in detail, tracking changes in purchases as affected by their marketing and CRM (customer relationship management) programs. Thus, retailers understand how they can grow the value of individual customers to their businesses.

In the e-tail world, when you click on an item or a page, a Weblog records what page you are on, what time you were on there, how long you spent, etc. An e-tailer can take those clicks and feed them into an engine. All the information is stored historically, so when another customer clicks somewhere, the engine will recognize the pattern and will know that it is appropriate to send a certain page to that person.

### *How can retailers use data mining to increase profits?*

Data mining can identify valuable customers who are likely to defect to a competitor, allowing the CRM team to target them for retention. It also points out potential long-term, high-value customers who can be accelerated to that value through marketing programs. Retailers can encourage the right purchase behavior. Retailers can make



marketing new products and services more profitable by using data mining to find customers most likely to respond to an offer for such products or services.

If people buy a certain basket of goods, you put one thing on sale in order to entice people to buy the other ones, because from the analysis you see that people tend to buy certain things together. Alternatively, you can place the goods physically close to each other.

### ***Are there different levels of data mining?***

Yes, but primarily from a technology and statistics perspective. Directed data mining allows users to specify what they are interested in discovering, such as finding good targets for a product. Undirected data mining uses a clustering approach that looks for pure statistical patterns that show why customers are like each other in any way, but often not in a business-focused way. A third set of techniques uses association of "basket" analysis to discover links between different products. This approach is not customer focused at all. The most famous example of this is when a super market spotted customers buying beer and diapers on certain days of the week. It was thought this was because men were making a diaper run after work and while they were there, picked up some beer. Seeing this pattern, the supermarkets placed the expensive beer right next to the diapers. This technique can be very useful for such product-focused wins, but can add even more value when such linked purchases are tied to the customer details. Fourth kind, visual data mining, is not just about raw statistics. Humans can do a lot of the mining by using visual and exploratory tools in conjunction with powerful statistical techniques.

### ***Is there a wrong way to data mine? Can a retailer have too much data?***

Too many companies leave data mining to the technical team, who often employ "black box" techniques that simply produce statistical patterns without regard for whether such patterns can be used in the business context. For example, one strong pattern discovered by a retailer using such tools was that daytime temperatures peaked three days after a peak in tomato sales. This information is useless unless the retailer was intending to move into weather forecasting.

However, a retailer cannot have too much data, as long as it is used correctly. Mistakes can be made with simple statistical tools if they are presented with too much data, but a more common problem is that most statistical and data mining tools just stop working when data volumes become very large. However, such restrictions should be in the past due to ever-increasing computer power at a relatively inexpensive cost. A lot of the new software technology can make very good use of all the additional storage and retrieval power.

It is better not to sample data. The idea of sampling is to take a portion of a retailer's database, and run analysis on that part. The problem is that by not analyzing all the data, a retailer may miss key data points that could form a relationship. I also think it is better to analyze the data directly within the relational database, as opposed to pulling it into



another engine. A retailer will leverage the parallel processing power of the database and therefore process faster. I don't think you can ever have too much data, but the more data you have the slower the analysis will be - but also the more accurate your results. It is important to use a scalable database with the ability to add more capability.

***Is there customer information available to retailers through the Internet that is not as easily available through other channels?***

E-tailers can use Weblogs to look back historically and see that other people who click on certain pages were interested in buying specific products. The Web allows you to do more real time analysis. Immediately when customers click information, it goes to the data mining engine. The engine can then take the information, digest it, and elicit a response. Based on click history it can display a targeted, personalized page for a customer and also make targeted offers. If the engine notes that a customer bought skis on a sporting goods page, the next time he clicks, the page that appears may be more geared toward skiing accessories. Or the e-tailer may direct him toward discounts for ski equipment. It's like having a really good salesman follow you around in a store, watching your every move.

***Do consumers feel safe giving retailers information?***

The key issue here is that the retailer must explain to the customer exactly what data is collected and what it is going to be used for. A case should be made for why it is in customers' best interests to provide their data, such as tailored service, products developed especially for them, or marketing campaigns targeted only where appropriate. The simple rule should be that a retailer should not sell customer data to third parties unless the customer has given their permission for that to happen. Even then, a retailer should consider carefully whether selling a customer's name to a competitor for 10 cents is really the best way to make money from that customer.

Give people the ability to not be tracked. Retailers can always give their customers incentives to provide them with information. They can offer loyalty cards, free gifts, or a chance to know how they will use the information, such as survey results.

***Is data mining for every retailer?***

Almost any retailer could gain some value from analyzing their data. The main driver for whether or not to do so will be the scale potential benefits and return on investment (ROI) compared to the cost of collecting, storing, and analyzing the data. Thus a specialist retailer, with very few products and customers, may gain little insight from data mining over and above their own knowledge of their business.

***What are the costs involved in data mining?***

There are both hardware and software costs. These costs typically scale with the size of the database. A significant part of any data mining initiative is also the consulting



services to set up the systems and applications in order to fit best with existing business practices, as well as to adapt these processes to a new way of operating. ROI can be measured in terms of increased revenues and profits, or in terms of cost savings, generated because marketing and CRM practices were specified by the data mining output rather than previous techniques. For example, if a marketing campaign gets a 2% increase in response through better target marketing, and a responding customer can be shown to be worth an extra \$200 per year, the overall ROI should be easy to calculate.