

Data Mining Application: Predictive Data Mining in Grocery Stores

Context

Data mining can yield massive advantages to corporate users within a retail context. However, grocery retail, perhaps more than most other retail operations, stands to benefit greatly from advances in data mining (and storage and access) technologies. This almost natural fit with data mining occurs for a number of reasons, which I will elucidate below.

Frequent, repeated interactions: Grocery purchases typically occur with some regularity. Even when the vast majority of customers (households) have the option of choosing between several stores, they typically go to no more than one or two retail locations for most of their general food shopping needs. Moreover, most large manufacturers of consumer packaged goods (CPG) will typically have products represented at both locations. Therefore, data can be captured regularly from customers' shopping behaviors.

Environmental variables that can be changed: Most large retailers do not keep the store environment the same for very long. In much the same spirit as Ian Ayres mentions Monster.com changing their advertising formats online, grocery stores will change prices, advertising, signage and store formats from week to week. Moreover, the additional featuring of items on weekly flier with different levels of discounts creates additional stimuli to which customers may or may not respond. The net result is an environment in which multiple variables over thousands of items (a typical grocery store can easily carry between 20,000 to 40,000 SKUs) are being changed all the time. Such an environment is ripe for experimentation.

New technologies: The technological aspects of retailing in a grocery store are changing. Store loyalty cards now enable regular, and relatively error-free capture of consumer purchase data at almost no marginal cost. Third-party companies such as Catalina now have the capability to issue product-specific coupons based on discounts provided by manufacturers and their own analysis of retailer loyalty card data. More recently, these companies have begun to develop their own experimentation regarding what sorts of offers and incentive campaigns perform best. They are able to do so by conducting precisely the sort of randomized trials of targeted offers that Ian Ayres mentions in *Supercrunchers*. By then comparing the sales metrics of targeted customers to a control group of households with the same characteristics (who do not receive any offer), Catalina is able to assess the effectiveness of the overall offer. Moreover, third parties are then able to sell this information and these findings to manufacturers and retailers who can use them to boost their own revenues and profits.

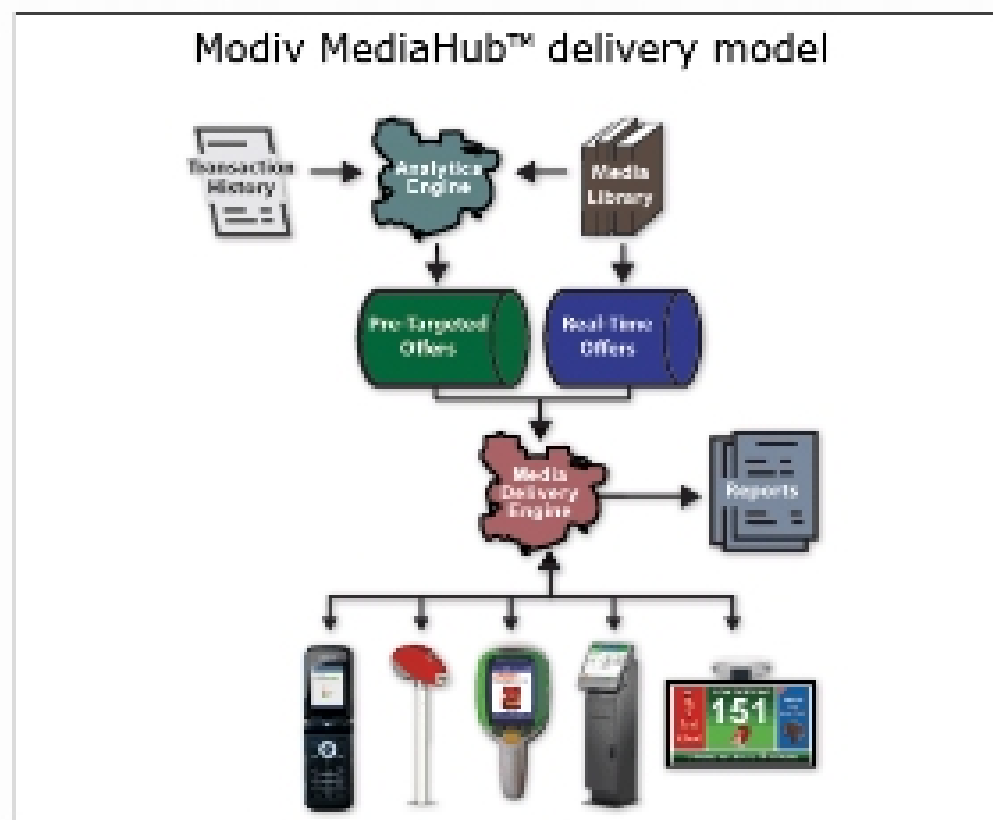
A "New" New Technology

All of the factors mentioned above, coordinated with data warehousing and management services, make data mining an extremely relevant pursuit in the grocery retail arena. However, an even newer technology has emerged that seeks to blend all of these data mining advantages that already exist in the retailing context. Companies such

as Modiv Media have developed proprietary hardware and software to enable customers to shop in different ways. A handheld scanner that allows consumers to “scan in” items as they pick them off the shelves is already being used at several Stop&Shop stores (the nearest one I have seen is in Hamden Plaza). Shoppers are already excited by this feature as it allows significant time-saving for what has been one of the traditional bottlenecks in the retail operation: the checkout line. If one could altogether bypass the checkout line, this would be a tremendous time saving for the shopper. As Sarah Shemkus points out in a Cape Cod Times article, the time and hassle saved by using this device makes it very attractive for the average, busy shopper. Moreover, presumably users would become more adept after several uses, making this an even greater time-saving device.

But if that were all there was to this technology, it would not be exciting from a data mining perspective! As the diagram below (from the Modiv Media website) shows, there are two-way arrows between the hand-held device and the IT systems that underlie the service. In fact, customers not only enter their purchase information, they are also allowed to receive (and immediately redeem!) coupons and other incentives provided by manufacturers or by the retailer. Moreover these devices also allow the retailer to track the customers’ location and progress over time.¹

Figure 1: Schematic Diagram of Modiv MediaHub
© Modiv Media



The confluence of all this information provides a whole new type of data mining problem: how should retailers choose which targeted incentives to give to customers in real time? In working on and refining the algorithms to solve this problem, companies like Modiv Media not only require significant data capture and warehousing capability,

¹ Another Pennsylvania-based company, Sorensen Associates, has independently developed a PathTracker service that allows retailers to target the position of shopper carts as they progress through the store. However, they do not currently detect what goes into/leaves the cart at any point in time.

but they also need to make full use of OLAP-like data-recovery, and complex (but quick to run) algorithms that help decide which incentives to issue, to whom and when.

While the specific design of these algorithms (and their progressive refinement) are highly proprietary and will likely be changing over time, the basic requirements would stay the same. In essence they are:

- (1) Do they get the customer to buy more/higher margin products as a result of the offers generated by the data mining?
- (2) Does the technology/do the offers convince the customer to buy more of my (the manufacturer funding the coupon or price drop) products?
- (3) How effective is it? What is the basic return on investment (ROI)? Who (the manufacturer or the retailer) receives the most benefit from funding such technologies and incentive programs?
- (4) What should change about the offer provided to the customer since last time an offer was provided?

Although not a direct requirement of the algorithm, the customer must also perceive this as an added service and not a complete nuisance that disrupts the natural flow of the shopping trip.

Despite the fact that much about these algorithms will never be made public, the very presence of these services indicates that some very sophisticated data mining is taking place as we stroll casually through the grocery aisles!

Sources:

Cape Code Times (<http://www.capecodonline.com/apps/pbcs.dll/article?AID=/20071120/BIZ/711200301>)

Modiv Media website

(<http://www.modivmedia.com/modivmedia/home.do?action=index>)

(<http://www.modivmedia.com/modivmedia/wrapper.jsp?action=hub>)

Catalina Marketing

(<http://www.catalinamarketing.com/>)

Sorensen Associates

(<http://www.sorensen-associates.com/whatwedo.html>)