

MIS and Marketing: Secrets of Strategic Information Mining

*By Boris Gendelev
Principal, Wheaton Group*

*Original version of an article appeared in the Summer, 1992 issue of
"Chief Information Officer Journal"*

[Note: Despite dramatic increases in raw computing power and a proliferation of end-user software tools since the publication of this article, virtually all of the content remains highly relevant. Interestingly, this article anticipated the wide acceptance of the Data Warehousing discipline by IT, and the creation of Data Warehousing departments.]

Data-based marketing is fairly new, so few CIOs have experience with the relevant methodology and technologies. Here's how to go from data processing to information mining.

One of the important challenges today's CIOs face is the shift from data processing to information processing. On the forefront of this phenomenon is perhaps the most strategic application of all: data-based marketing. At the core of data-based marketing is the mining of historical transactional data to uncover customer patterns and trends.

Data-based marketing cannot succeed without support from technology experts. Unfortunately, marketers often find IS personnel uncooperative. The problems usually stem from some basic misconceptions:

Misconception: The MIS department has the knowledge and tools to build correct data-based marketing systems; it just needs to move more quickly and pro-actively.

Reality: MIS's experience base is usually operational systems. An order-entry clerk's very regimented use of data does not resemble the way marketers use information to devise customer-acquisition strategies, plan promotions, and search for new marketing ideas. Thus, most of what IS personnel learn from building operational data processing systems simply doesn't apply to data-based marketing.

Misconception: Marketers do not communicate what they want.

Reality: Marketing requirements differ significantly from other business requirements. Marketers cannot communicate a complete and invariant set of requirements because their most important requirement is to be able to deal with constantly changing needs.

Misconception: The way the data already exists in the operational databases is good for marketing information mining.

Reality: For marketing needs, the data must be carefully prepared to address ever-present integrity and consistency problems. Moreover, the data must be cast into logical and physical structures tailored to the unique task of marketing information mining. Resource sharing between operational and informational databases usually leads to bottlenecks and escalating costs.

Misconception: Relational queries give users enough flexibility for accessing the data.

Reality: Relational interfaces cannot do complex data transformation and statistical aggregation in a straightforward and efficient way. Expressing marketing analysis queries in SQL is about as natural as writing operating systems in COBOL. This is the reason that, in the absence of their own database, marketing analysts may use SQL to pull data extracts, but they do the real work with other tools.

Misconception: End-user "automated" analysis tools, based on rule induction, neural networks, fuzzy logic, genetic algorithms, fractals, or fuzzy logic, replace the need for human information mining.

Reality: All these techniques require, just as old-fashioned statistical analysis does, careful structuring of the inputs and tinkering with the knobs. At the very least, a human analyst must discover what is relevant before asking a program to verify, refine, and quantify it.

Misconception: Data-based marketing is just a sales forecasting or a customer-selection system.

Reality: Analyzing marketing data and implementing the results of the analysis are two different things. Information mining will likely result in a slew of new operational systems, but one should not confuse gold with the process of mining it.

Because data-based marketing is new, few CIOs have experience with the relevant methodology and technologies. CIOs must understand the key differences between data processing and information mining. The goal of data processing is to support the smooth flow of a business's daily activities. The goal of information mining is to detect and measure marketplace phenomena in order to actively manage business change.

Different Process

Because of differences in purpose, data processing and information mining use computers in very different ways. Information mining is characterized by the use of:

- Long, detailed histories of interactions with each and every customer, as opposed to just current or highly pre-summarized data.
- Data dynamically derived from the basic elements by computations, re-coding, etc., rather than stored static data.
- Statistical aggregation of data rather than retrieval of individual record values.
- Ad hoc, data-driven iterative processing rather than a well-defined flow of execution steps.

- Individual project work organization.

These characteristics lead to wide swings of resource utilization, greater need for resource flexibility, and low reuse rate (and therefore little opportunity for traditional systems quality assurance).

Information mining is done not through a collection of well-specified applications, but in a computational environment that facilitates data-intensive research.

Methods and Technology

A handful of basic concepts provide the foundation for a good information-mining architecture:

- Support for the Time Slice, Classify. Measure, Analyze, and Model cycle.
- Customer-centered data organization.
- Dedicated computing resources.
- Availability of slack resources.
- Focus on contents of the data.
- Focus on result verification.
- Support for core marketing DSS (decision support systems) and EIS applications.

Although unpredictable in each specific instance, information mining has patterns. Therefore, you can build specialized software to facilitate it. One might get an impression, particularly after seeing packages that claim to support data-based marketing, that the process is very simple:

- Classify each customer based on the current data and summarized history. For example, one classification might be based on gender, another on life-to-date number of orders.
- Use conjunctions to isolate customer segments – for example, female customers with more than two orders, male customers with more than one order.
- Select and count customers in the segments of interest.

This type of processing supports marketing based on a mix of intuition and primary research, such as surveys and focus groups. One must rely on intuition and primary research for launching tests of new products and promotions because there is no customer history one can use to determine analytically who the best candidates are. Cross-tabulating customers by various characteristics helps marketers understand who their customers are and even allows them to define customer clusters, which are mini-markets around which marketing programs are developed.

But the biggest payoff of a marketing database comes from the ability to practice analytical (data-driven) marketing. The analytical cycle is more complex:

- Categorize each customer, as of a chosen historical slice of time, using any or all of the available data. For example, how often did he buy, and in what range was his average order in 1990?

- Measure what each customer did after that point, or how each would be categorized in the next time slice. This can be numeric measures such as dollar purchases in 1991, or descriptive categories such as average order range in 1991, or both; for example, 1991 purchases by product category.
- Summarize numeric measures or tabulate descriptive ones by categories across all customers.
- Analyze summaries looking for differences in measures between categories. For example, how much better do customers with high-average order in one year do in the following year, compared with low-average order buyers?
- If needed, create a numerical model of the discovered relationships. For example, one model might be:

Expected customer spending next year = ((average order x 2) x (prior number of purchases)) / number of months since the last order.

This cycle appears in just about every marketing analysis:

- Before vs. after analysis: By how much and where did a competitor's price drop affect the company's business?
- Test vs. control: Is the new creative package better than the old one?
- "What-if" scenario evaluation: How much more profitable would our marketing campaigns be if customers were segmented differently?
- Analysis of customer potential: How profitable is each customer segment over a 10-year period? How do we acquire more profitable customers?
- Affinity analysis: Can we profitably specialize merchandise and promotions?

Good software is the key to making this processing cycle effective. Summarizing numeric data by categories is familiar to MIS, for even data processing systems have some management reports. The process of analyzing summaries is relatively easy using spreadsheets or other EIS IDSS software such as IRI's Express. Model construction can be done with statistical software such as SAS or SPSS, or with rule induction or neural learning packages. However, the process of creating complex, ad hoc categories and measurements has not yet been widely addressed. The needed capabilities would be best provided by a hybrid package that combines:

- Storage and scanning efficiency of sequential master files.
- Transparency of primary relationships found in hierarchical and network models:
- Relational flexibility of linking data ad-hoc.

- Object-oriented qualities of polymorphism and inheritance for hiding differences and sharing commonalities of definitions.
- Numerical transformation and aggregation capabilities of statistical packages.
- Support for user-created function libraries.
- Easy integration with DSS/EIS software.

Going Back to the Future

By examining how marketers use historical marketing data, the CIO discovers the following:

- There are hardly any individual record queries; the method of information mining is wholesale aggregation.
- Most marketing concepts are represented not by existing fields, but by ad hoc computed quantities and categories – even if marketers are using a concept over and over again they may apply it to different time frames.
- Almost always, marketers process all the data pertaining to each customer together because the most interesting elements are at the lowest level of the logical hierarchy (such as order line items).
- Most of the time marketers need to process all customers in order to compare customer segments, not just get information about one group.

Not surprisingly, for this kind of information mining, a master-file approach has proven superior to other database organizations because:

- Separate tables with foreign keys do not offer any advantage and could carry a tremendous overhead – on the other hand, when all customer data is already physically together, storage is much smaller and scanning is much faster.
- Indexing schemes are of little use because they rely on static views of data and assume that small subsets are processed at a time – instead, the secret to high efficiency is making all dynamic calculation memory-based.
- File organization as such doesn't stand in the way of having high-level, non-procedural access tools and flexibility of adding data elements – with object-oriented technology such capabilities are not hard to create for any file structure.

Dedicating Storage and Processors to Information Mining

The integrated world of MIS often considers segregating databases and creating data redundancy a capital offense. But, as Inmon observed, not doing so may lead to much greater and uncontrollable

redundancy, with every user pulling his own extracts to get his job done. A separate historical database, (or in Inmon's words a "data warehouse"), minimizes and controls redundancy.

Having processors and storage dedicated to information mining avoids the conflict that arises if you introduce erratic information processing into an environment of predictable utilization rates. Fortunately, unless your customer file contains the entire population of the United States and all citizens' purchases, you may not need very complicated and costly hardware.

Once all parties agree to separate computing resources, periodic, not continuous, feeding of data from operational databases is a natural outcome. The strategy of updating the marketing database only periodically has few drawbacks and several important advantages:

- It permits creation of a Data Quality Filter (discussed later) to assure data usability.
- Iterative analysis is best done on data that are not changing.
- Continuous updating takes up resources needed for data analysis. Periodic updating fits well with peaks and troughs of information mining.
- Not having the most current layer of data can be easily compensated by straightforward short-term projection of customer counts. Most of the time, it is not even an issue because analysis is done by time slicing the past.
- Short-term promotion tracking reports can be easily produced from the operational databases.

Resources Ready to Deal with Peak Demand

Dedicating resources to information mining is not enough. Dealing with peaks of demand that information mining creates requires either having much more hardware than the average demand, or being prepared to upgrade on short notice. Information mining and analysis can be done only in concentrated efforts. And the need for mining usually increases. These and the following considerations suggest that microprocessor-based workstations should be the platform for marketing's departmental information mining:

- Availability of large arrays of RAM at very low cost (less than \$50 per megabyte).
- Surprisingly fast processors; already faster than most mini-computers.
- Inexpensive, very large capacity, expandable (up to tens of gigabytes, at less than \$2.00 per megabyte) disk storage, with throughputs up to 120 megabytes a minute.
- Availability and affordability of end-user and software development packages.
- Ease with which hardware can be upgraded and reconfigured, all under an analyst's control.
- Phenomenal curve of improvements in speed and capacity.

Focus on Data Content

Having sufficient resources available allows marketers to concentrate on marketing data, but only if the data is usable. To focus on data content, CIOs should not view the files and fields as just

placeholders for passing data, as it is in data processing. The essence of an informational database is not its structure but its content.

To take good care of data, place a Data Quality Filter between operational and informational databases. This filter uses common sense as well as statistics to assure that:

- Every piece of incoming data is audited for accuracy, completeness, and consistency with already-accumulated data – deficiencies are returned to the feeders.
- Data is standardized and reorganized around the most likely subjects of analysis – customer households or business organizations.
- Changes are logged for data without operationally maintained histories.

In a large, vertically integrated company with many data feeders, a Data Quality Filter might be part of a centralized data repository serving several different departments. There are products that automate the process of construction and maintenance of large company-wide data warehouses. They feature:

- Generation of COBOL/SQL and JCL code for pulling data from operational systems and transforming it to achieve consistency and accessibility.
- Maintenance of time-dependent data.
- Transparency of DMBS technology for both source and target databases.
- Active dictionary of mappings of stored data to its sources and between different levels of warehoused data.
- Template database models for many businesses.

Avoiding Disaster: Verifying Results

Having a high-level, descriptive specification language or visual interface goes a long way toward assuring quality results, particularly if the data passed through a Data Quality Filter. However, given the strategic nature of information mining, that is not enough. Marketing analysts must verify each and every result produced. Certain features of the environment would promote quality control:

- Support for sampling of data and one-pass execution of several requests, which removes processing delay penalty for incremental specification.
- Maintenance of all prior results in a library such that all results related to a given concept can be retrieved.
- Support for incremental development of concepts such as a concept dictionary with project and time partitioning.
- Easy access to information about the origins of each data element and the caveats of its use.

One must also be careful in moving data processing personnel into the role of marketing information providers. The habits of focusing on the structure of files and the logic of programs instead of the

content of data and the meaning of the outputs might be hard to overcome. It is particularly difficult if the same people still have to play data processing roles.

From Information Mining to Applications

Certainly not all information-mining efforts lead to the creation of new applications. Some do not even produce interesting results, let alone influence strategies or tactics. However, the most common applications that emerge are:

- A customer-acquisition planning system that helps marketers choose the best ways to acquire new customers based on models that project the long-term payoff of such efforts.
- A promotion planning, customer selection, and tracking system based on a segmentation model that ranks customers based on expected profitability – a financial model combined with a model of customer long-term value determines the depth of selection for targeted promotions.
- Tracking and projection of critical customer segments – this is an EIS application used to keep a watch on the "health" of a customer base, project sales, and play "what if" scenarios with the marketing strategy.
- A test planning and evaluation system supported by well-defined customer clusters.
- Merchandising support based on discovered clusters of products that customers tend to buy as a group.

The use of these systems leads to new ideas and new research questions that translate into more information mining. CIOs should develop and execute these marketing and executive applications in the information-mining environment for the following reasons:

- In the operational environment it will be difficult to get data of the same quality and consistency as in the historical informational data-base.
- Moreover, although these applications are not as fluid as information mining itself, they need to be considerably more open to revisions than order entry or accounting.
- A compelling argument for maintaining these applications within the information-mining environment is that quality-control procedures established there are more appropriate than regular data processing quality controls.
- A crucial element in executive information systems is a human information provider, usually a marketing data analyst. Information providers perform information mining, investigate suspicious results, and answer follow-up questions. The place for these is the information-mining environment.

MIS' Role and Opportunities

The CIO must know the limitations of the MIS department's current methods and technologies and the personnel trained in them. Companies that had to provide information-mining services have learned, to various degrees, more appropriate techniques.

Some IS organizations, after considering the issues of information mining, will decide to just stay out of the marketers' way, although they cannot avoid dealing with some data-collection issues. In these situations marketing may seek outside help.

At the same time, the most restless members of the IS organizations may welcome the challenge and embrace the opportunity to achieve high-level visibility. When this is the case, the need for outside help may be only short term. Others may even want to take over the function of information providers. Their success will depend on how well they have mastered the fundamentals of data-based marketing support.

Boris Gendelev is a Principal at Wheaton Group, and can be reached at 847-205-0916 or boris.gendelev@wheatongroup.com. The firm specializes in direct marketing consulting and data mining, data quality assessment and assurance, and the delivery of cost-effective data warehouses and marts.