# Online marketing research

A. Agrawal
J. Basak
V. Jain
R. Kothari
M. Kumar
P. A. Mittal
N. Modani
K. Ravikumar
Y. Sabharwal
R. Sureka

*Marketing decisions are typically made on the basis of research conducted using direct mailings, mall intercepts, telephone interviews, focused group discussion, and the like. These methods of marketing research can be time-consuming and expensive, and can require a large amount of effort to ensure accurate results. This paper presents a novel approach for conducting online marketing research based on several concepts such as active learning, matched control and experimental groups, and implicit and explicit experiments. These concepts, along with the opportunity provided by the increasing numbers of online shoppers, enable rapid, systematic, and cost-effective marketing research.*

## 1. Introduction

Estimating the relationship between marketing and response variables is fundamental to marketing- and merchandizing-related business decisions. Consider a simple example in which a retailer must select the price at which to sell a certain item. A systematic decision requires the retailer to know the relationship between the price of the item (*the marketing variable*) and the demand for the item (*the response variable*) at the various price points.

As a (slightly) more complex example, consider a situation in which the retailer feels that running a promotion on an item will lead to increased overall revenue. The promotion may take the form of a temporary price reduction achieved through the use of a coupon. Setting the face value of the coupon determines the effective price at which the item is sold, and this can be determined only if the demand at various price points is known. However, the decision is more complex if one considers other effects. If the retailer sells multiple brands of the item, reducing the price of a particular brand may result in *shifting* the sales from a competing brand to the promoted brand, leading to flat overall revenue. Also, shoppers may stock up on the item during the promotion period, leading to reduced sales of the item following the promotion period and net flat revenues.

Though simple, these examples illustrate the complexity of marketing and merchandizing. One may pose the problem so as to be amenable to analytical techniques by saying that an *informed* marketing and merchandizing decision requires estimating the multivariate relationship between marketing and response variables. Put simply, it involves knowing how the response variable(s) will change when one or more marketing variables are changed.

Estimating the behavior of a response variable to a change in the marketing variable requires data. Typically, data is collected through marketing research conducted through direct mailings, mall intercepts, telephone interviews, focused group discussion, and the like. In the simple example considered above, through telephone interviews one may simply ask the consumers to indicate the likelihood of their buying the item at different price points and use the collected data to infer the relationship between the marketing variable of interest (price) and the response variable (demand). The one-on-one interaction required in some of these modalities of collecting data (for example, in telephone interviews) coupled with the large turnaround time (for example, due to the transit time of a direct mailing to and from the respondent) and the significant number of person-hours required renders this traditional form of marketing research expensive, slow, and susceptible to inaccuracies.

The rapid growth of the Internet creates an opportunity for conducting online marketing research (OMR). Indeed, by some estimates, about 60% of the population of the United States and the European Union has Internet access. Collectively, these regions also account for a substantial amount of the world purchasing power according to the British Market Research Association (BMRA) [1] and the World Association of Opinion and Market Research Professionals (ESOMAR) [2]. Separately, various regions in Asia are also showing signs of increased Internet access. This widespread adoption of

©**Copyright** 2004 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.



0018-8646/04/$5.00 © 2004 IBM



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the Internet makes a large cross section of the population accessible through the Internet and ensures that the needs and preferences of a substantial and representative population of the consumers can be obtained online.

This paper is motivated by the possibility of providing actionable business intelligence rapidly, systematically, and cost-effectively through OMR. Given the complexity of marketing and merchandizing decisions in modern businesses and the limitations of space that are necessarily enforced by the paper format, we have chosen to focus on some fundamental aspects of OMR. Specifically, we focus on those aspects that will benefit any serious attempt at realizing an online marketing research implementation.

We have organized the rest of the paper as follows. In Section 2, we provide a conceptual overview of a system and describe a basic setup that can be used to conduct OMR. Our focus is not on system-level internals, since these are dependent on the commercial server in which OMR is implemented. Rather, we seek to provide some idea of the chain of events that occur, the various control points for OMR, and the various innovations proposed in this paper. These innovations, we believe, are central to OMR, and we detail them in Section 3. In Section 4, we present an overview of some types of actionable business intelligence that can be obtained on the basis of the proposed system and the algorithms. We conclude in Section 5 with some discussion.

## 2. Schematic for implementing online marketing research

It is useful to distinguish between OMR and traditional business analytics. Business analytics uses existing data, perhaps collected during normal business operations, to find the relationships among variables (including marketing and response variables). In contrast, marketing research *intentionally changes the marketing variable* and collects the corresponding response. Data obtained from marketing research is thus more suitable for establishing the relationship between specified marketing and response variables. Consider, for example, that a manufacturer wants to know, as part of the exercise of designing a new product, the value (utility) that buyers associate with the different features of the product. Such information allows the manufacturer to implement desired features in the new product and remove the features that do not have utility for the buyers. Business analytics would attempt to uncover the utility of the features from historical sales of perhaps different models that the manufacturer has sold. However, since there was a certain feature that was always present, there is no way to establish whether the absence of that feature would affect sales. On the other hand, marketing research would change the features and attempt to determine the likelihood of shoppers buying the contrived product (for example, with a survey in which users indicate the likelihood of buying one of several contrived products). Though there are systematic methods of arriving at contrived products [3–5], at this point we simply wish to highlight that the data collected under the proactive change in the marketing variable would be more suitable for establishing the relationship between the marketing (features of the product) and response variable (sales).

In OMR, the change in the marketing variable is done online, and the responses are also collected online. Since the process of changing the marketing variable and collecting the responses cannot interfere with the normal operation of the site, we first describe the site and highlight how the change in the marketing variable is achieved. We assume that there is a Web site whose pages comprise content typical of online sites. In particular, there are some navigation controls, there is some space for horizontal and/or vertical banner advertisements, and there is space for the main content of the page. Hyperlinks or "hot spots" embedded within the main content or navigation controls allow a visitor to navigate the site and engage in transactions offered by the site. The content of the banner advertisements is chosen by logic embodied in the *recommender subsystem* of the commerce server on which the online site is developed. Certain activities (for example, a purchase) require the individual to log in, while other activities do not require identification to the system. For the latter, the individual can browse the site as an anonymous user.

The marketing variable is changed through the horizontal or vertical banner advertisements. For example, when price is the marketing variable, a coupon can be shown to the user in a horizontal or vertical banner [6]. The coupon changes the effective price of the item for the shopper (change in the marketing variable), and the user's acceptance of the coupon (by clicking on it) and subsequent redemption is the response variable. There are additional mechanisms that may be used to change the marketing variable, though here we assume that all marketing variables are changed by changing the content of the horizontal or vertical banners. This does not in any way reduce the generality of the proposed approach and helps in making the discussion clearer.

**Figure 1** provides an overview of the chain of events. The overall flow begins with the specification of an objective on the part of the merchant—for example, "determine the demand as a function of price," or "find the effect on the sales of Brand B caused by a discount on Brand A," and so on. The marketing and response variables are identified, and a data-gathering activity is initiated. We call each such data-gathering activity an "experiment"; the deployment of an OMR experiment may utilize other subsystems of a commerce server. The OMR experiment then changes the marketing variable



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

for selected visitors to the Web site and assigns each selected visitor to a group to create matched control and experimental groups (explained in greater detail in Section 3). The visitors (users) can be selected on the basis of their clickstream (navigation pattern), along with the user's historical transactions (if the user is logged in) and other information. Nonetheless, the use of the clickstream provides a way of selecting a user even when the user is simply browsing as an anonymous user without actually logging in.

The response of each selected user is measured either explicitly or implicitly. The experiment can execute for a prespecified time period, or an information-theoretic criterion can be used to terminate the experiment when the gain in information from additional data collection falls below a certain threshold.

For example, the control group may not see anything related to the product, and there may be three matched groups which are offered e-coupons with a discount value of 5%, 10%, or 20%. The differential response of the groups then provides a basis on which to extrapolate the demand at various price points.

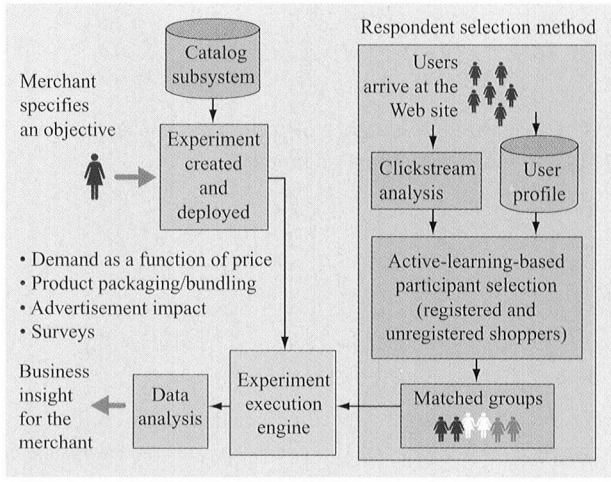



**Figure 1**

Overview of the online marketing research solution.

## 3. Foundations of systematic online marketing research

Several innovative features required to make OMR systematic, rapid, and cost-effective are described in the sections that follow.

### *Matched control and experimental groups*

A significant part of OMR is driven by observing the change in response resulting from a change in the marketing variable (price, bundling of products, and so on). This requires that the possible effect of any variable other than the changing marketing variable be removed. OMR achieves this through the use of matched control and experimental groups. To illustrate how this increases the accuracy of the inferences made, suppose that a merchant wishes to ascertain the demand at various price points. Say, multiple groups of customers are formed on the basis of random selection, and each group is offered the product at a certain price. The difference in the overall response of the various groups cannot be attributed entirely to the change in price (unless the sample size of each group is large). This is because differences in population characteristics also contribute to the difference in the response of the groups.

OMR thus uses the concept of matched control and experimental groups. It chooses a potential respondent using active learning (see below) and assigns the individual to the control group. Each experimental group is then assigned a unique individual who matches the individual in the control group on the basis of a specified set of user attributes within a specified tolerance. The attributes used in deciding the degree of match between the two individuals include demographics-based information, session-based information (such as shopping cart total), and clickstream information. The clickstream-derived matching criteria can be specified as follows: "Users A and B are similar if they have individually visited pages p1, p2, $\cdots$, p$n$" and allows matching of users who may not be registered or who have not logged in. More systematically, we have developed a method for finding the "distance" between two clickstreams [7]. Our method is based on estimating the distance between two pages; in theory, the distance between two pages should be based on semantic analysis of the page contents. However, most Web pages contain images (or other multimedia-based data), and the present state of technology does not allow such a semantic analysis. Our method thus uses the joint probability of occurrence of two pages to estimate the distance between them. Our rationale is as follows: If users (on a statistical basis) visit a page (say, A) and then visit another page (say, B), there must be a strong content-based connection (and hence similarity) between the two pages. Mathematically, say there are $C$ clickstreams. Denote the sequence of pages in the $i$th clickstream as $[(A_{i1}, t_{i1}), (A_{i2}, t_{i2}), \cdots, (A_{in}, t_{in})]$, where the first subscript denotes the clickstream number and the second subscript denotes the sequence in which the page was visited during that session. The symbol $A$ is used to represent a page, and the symbol $t$ is used to denote the time at which the page is accessed. The joint probability of occurrence of two pages, say $A_m$ and $A_n$, can then be defined as



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

$$P(A_m, A_n) = \frac{1}{T} \sum_{j=1}^{C} \sum_{k=1}^{|A_j|} \sum_{l=1}^{|A_j|} I[(A_{jk}, t_{jk}), (A_{jl}, t_{jl})],$$

where $T$ is the total number of page pairs considered and $I[\cdot]$ is an indicator variable defined as

$$I[(A_{jk}, t_{jk}), (A_{jl}, t_{jl})] = \begin{cases} 1 \text{ if } \begin{cases} \|k - l\| < \theta_g, \\ \|t_{j,k+1} - t_{jk}\| > \theta_v, \\ \|t_{j,l+1} - t_{jl}\| > \theta_v, \\ \|t_{jl} - t_{jk}\| < \theta_e, \\ A_m = A_{jk} \text{ and } A_n = A_{kl}; \end{cases} \\ 0 \text{ otherwise,} \end{cases}$$

where $\theta_g$, $\theta_v$, and $\theta_e$ are specified constants. The indicator variable evaluates to 1 if the pages are accessed within a certain distance of each other ("gap"); if the access times between successive pages are greater than a threshold ($\theta_v$); and if the individual pages are viewed for a reasonable period of time ($\theta_e$). These restrictions ensure that crawler-based actions are excluded and that sessions which have interruptions are not considered to be co-occurring. The indicator function is thus designed such that pages which *actually* occur close to each other (in terms of space and in terms of time) contribute to the joint probability of occurrence. Distance between the pages can then be a monotonically decreasing function of the joint probability of occurrence [for example, $1 - P(\cdot)$]. The cost of transforming one clickstream into another using insertion, deletion, and replacement of page views is then taken as the distance between the clickstreams (for additional details, see [7]). We have designed a simulator which mimics a Web site and simulates users with known interests and preferences in order to generate the data to test our algorithm. The contents of the pages are known in this simulated environment, and our comparison shows that the proposed method comes close to approximating the distance derived from semantic analysis. Of course, semantic analysis is not possible in a real setting, and the proposed method can be used as an accurate estimate of the true distances between pages and the results used to compute whether a user matches another user on the basis of their individual clickstreams.

Using the above strategy ensures that the groups are "matched" in the sense that for each user in a group, a similar user exists in each of the other groups. A marketing variable can thus be changed between the groups, and the effect of the change can be measured directly from the difference in responses of the groups (the groups are similar to each other, with the single exception that they are exposed to a different marketing variable). Clearly, it is more difficult to do this matching with the more traditional forms of marketing research, in which implicit information such as clickstream is not readily available.

### *Active learning*

In order to conduct OMR rapidly and to limit the exposure of an experiment to the smallest possible subset of users, it is necessary to choose the respondents with care. Conceptually, the most informative participants should be chosen in such a way that it is possible to collect the required data using the minimum number of respondents. Learning from chosen participants (or "data points" in a generic context) as opposed to learning from the available data (or randomly sampled data) is often called *active learning* and has been the object of sustained study [8–12]. Typically, one begins with a small set of labeled data points (previous participants whose responses are known) to find the unlabeled data point (the next visitor to the site) which, if labeled (chosen as a participant), would provide the maximal gain in information. In the present context, one may begin with a few users whose behavior is known (through observation or through manual curation) and use an algorithm to find a user whose responses to an OMR experiment would be maximally informative. Technically, the prior approaches to active learning have been based on using the known (or labeled) data to find the next most informative data point. We have developed an innovative algorithm that actually reverses the role of the unlabeled and labeled data [13] and that uses available information such as demographics and clickstream to evaluate the anticipated gain in information that would result from the individual's response. Informative individuals are chosen for participation in the online marketing research experiment.

To clarify, let the attributes derived from demographics, clickstream, and historical transactions be denoted by the vector $x$ and the total information provided by an individual be denoted by $I(x|X)$, where $X$ represents the individuals who have already been sampled. Then, the next most informative respondent satisfies the relation

$$\underset{x, x \notin X}{\operatorname{argmax}} [I(x|X)].$$

It is possible that the most informative visitor, as determined by the above equation, may in fact never arrive during the course of the experiment. We thus recommend discretizing the entire feature space and computing the information content of the features in each feature cell. Each feature cell corresponds to an idealized user, and the most informative feature cells provide the set of most informative users. Any real user visiting the site and matching anyone from the set of informative idealized users can be selected as a potential respondent. If the set of idealized users chosen is large, it ensures that informative users are not discarded simply because they are not the *most* informative users.

To associate the information content corresponding to a certain feature vector (user), we form multiple models



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

that predict the behavior of a user given $x$. The notion of entropy (degree of disagreement) between these multiple models is then used to characterize the gain in information that is likely to result from an individual's response. Additional details of the algorithms are available elsewhere [8]. One may observe that the true behavior of a user is not required in this evaluation—instead, the degree of relative disagreement between the models is used.

The net result of active learning is that each of the groups formed is of compact size, relieving the downstream processing load and reducing the total time required to obtain business intelligence. When an incentive is offered to the participants (such as an e-coupon or a discount on a future purchase, say for participation in a survey), active learning also minimizes the total amount of expense incurred (in terms of the cumulative total of the discounts). Further, it minimizes the number of users that are exposed to change in the marketing variable. This localization ensures that OMR can be conducted with minimal impact to the normal operation of the site.

### *Implicit and explicit experiments*

Implicit experiments do not disturb the normal shopper flow and rely on the observed response to a change in a marketing variable for inferring the relationship between the marketing and response variables. On the other hand, explicit experiments disturb normal shopper flow and require the explicit participation of the shopper. Consider, for example, the task of estimating the demand as a function of price. An implicit experiment may create multiple matched groups and expose each matched group to a different price (by offering coupons of different face value to each matched group). The difference in the response (user acceptance of the coupon and subsequent redemption) can then be used to construct the relationship between price and demand. An explicit experiment, on the other hand, can be based on a survey in which the users are asked to indicate the likelihood of their purchasing the product at different price points. Implicit experiments are less distracting and often more accurate, since they do not make the shopper conscious of a question being asked and have a greater probability of capturing the shopper's true intent. To the greatest extent possible, OMR should use implicit experiments.

These key innovations serve as cornerstones for systematic, rapid, and accurate online marketing research. Clearly, aspects such as matched groups are difficult to create in traditional forms of marketing research, but they can be constructed online, thus improving accuracy. Similarly, the use of implicit experiments (to the greatest extent possible) ensures greater accuracy, while the use of active learning minimizes the cost (especially if a coupon or other price-reduction mechanisms are used) and increases the speed. Besides enabling marketing research for businesses with budgetary constraints, online marketing research provides an opportunity for continual adaptation of the operational and strategic aspects of business to enterprises as well as small and medium-sized businesses.

## 4. Example of actionable business intelligence from OMR

The concepts presented in the previous sections are surprisingly powerful in the range of actionable business intelligence that can be provided to a merchant. We provide a small sampling of the possibilities:

- *Determining price sensitivity:* The price sensitivity of a product can be measured with matched groups, with each group being offered a variable discount based on offering e-coupons of varying face value to the individual groups. The response can be used to approximate the (unknown) relationship between price and demand. Segment-specific price sensitivity can be similarly determined, with all of the individuals in each group being restricted to the specific segment for which the price sensitivity is desired.
- *Determining cannibalization effects/brand loyalty:* Often a discount on an item increases the sales volume of that item at the expense of the sales volume of other items. One way of estimating the cannibalization effect is to select matched groups who have the product in question in their shopping cart. To each matched group except the control group, discounts of increasing amounts are offered on a competing product. The number of individuals who abandon the original item coupled with the discount value at which the switching occurs provides insights into brand loyalty.
- *Catalog reordering:* Product displays are known to have a correlation with sales [13]. Products that must be promoted are often displayed more prominently. By observing the response of matched groups to different display sequences, it is possible to extrapolate a sequence that is optimal for a given online store.
- *Deriving attribute utilities:* By constructing orthogonal arrays [3–5], it is possible to see the differential response of the matched groups to products which differ in only a few of their attributes. The utility of each attribute can then be ascertained.

We have created a proof-of-concept prototype of these and other forms of OMR on top of the WebSphere* Commerce 5.4 BE server. Since OMR is not a part of the product, it is not possible to quantify the benefits that can result from its use. However, it seems reasonable to assume that such functionality can facilitate informed



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

marketing and merchandizing decisions. At a more rudimentary level, many Web sites do use an explicit form of marketing research. For example, pop-up windows asking a single question or information pages which have a question ("Did you find this information useful?") represent the first step in online marketing research. However, we are not aware of systematic methods for online marketing research such as those proposed in this paper.

## 5. Discussion and conclusion

The rapid growth of the Internet as a medium for commerce prompts the need for continued innovations. Virtual businesses, formed on the basis of partnerships of smaller businesses and other changes, are redefining the requirements for future-generation commerce servers.

However, there is little doubt concerning the need for a business to know its customer base. As is known, retail strategies and customer characteristics have a great impact on the response of promotions [14]. From that perspective, marketing research has substantial potential for allowing a business to know its customers better.

Privacy concerns may arise from any online marketing research activity. Various associations [15] provide guidelines on the appropriate use of the Internet for opinion and marketing research. When conducted appropriately, OMR benefits businesses of all sizes, increases competition, and delivers increased value to the customer. Indeed, the offering of an e-coupon based on gathered intelligence allows a merchant to make relevant offers to customers while at the same time allowing customers to increase their buying power.

*Trademark or registered trademark of International Business Machines Corporation.

## References


1. British Market Research Association; see *http://www.bmra.org.uk/*.
2. World Association of Opinion and Marketing Research Professionals; see *http://www.esomar.org/*.
3. C.-X. Ma, K.-T. Fang, and E. Liski, "A New Approach in Constructing Orthogonal and Nearly Orthogonal Arrays," *Metrika* **50,** 255–268 (2000).
4. A. S. Hedayat, N. Sloane, and J. Stufken, *Orthogonal Arrays: Theory and Applications*, Springer-Verlag, New York, 1999.
5. R. D. Cook and C. J. Nachtsheim, "A Comparison of Algorithms for Constructing Exact D-Optimal Designs," *Technometrics* **22,** 315–324 (1980).
6. M. Kumar, A. Rangachari, A. Jhingran, and R. Mohan, "Sales Promotions on the Internet," *Proceedings of the Third Usenix Workshop on Electronic Commerce*, 1998, pp. 167–176.
7. R. Kothari, P. Mittal, V. Jain, and M. Mohania, "On Using Page Co-occurrences for Computing Clickstream Similarity," *Proceedings of the SIAM International Conference on Data Mining*, 2003, pp. 154–165.
8. V. V. Fedorov, *Theory of Optimal Experiments*, Academic Press, Inc., New York, 1972.
9. D. J. C. Mackay, "Information-Based Objective Functions for Active Data Selection," *Neural Computation* **4,** 589–603 (1992).
10. V. Iyengar, C. Apte, and T. Zhang, "Active Learning Using Adaptive Resampling," *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2000, pp. 92–98.
11. N. Abe and A. Nakamura, "Learning to Optimally Schedule Internet Banner Advertisements," *Proceedings of the 16th International Conference on Machine Learning*, 1999, pp. 12–22.
12. R. Kothari and V. Jain, "Learning from Labeled and Unlabeled Data Using a Minimal Number of Queries," *IEEE Trans. Neural Networks* **14,** No. 6, 1496–1505 (2003).
13. J. Coggins and B. Senauer, "Grocery Retailing," in *U.S. Industry in 2000: Studies in Competitive Performance*, D. C. Mowery, Ed., National Research Council, National Academic Press, 1999, pp. 155–178.
14. S. Dhar and P. Rossi, "Why Responsiveness to Retail Promotions Varies Across Retailers," *Research Highlights from The University of Chicago Graduate School of Business*, Vol. 4, No. 1, 2002; see *http://gsbwww.uchicago.edu/news/capideas/summer02/retailpromotions.html*.
15. *Use of the Internet for Conducting Opinion and Market Research: Ethical Guidelines*, Market Research Association, Rocky Hill, CT, 2000.


*Received October 15, 2003; accepted for publication February 27, 2004*



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

**Aseem Agrawal** *The World Bank, Washington, District of Columbia 20433 (aagrawal@worldbank.org).* Mr. Agrawal graduated from the Indian Institute of Technology in Delhi and received a master's degree from the Indian Institute of Management in Calcutta. He has worked in the fields of finance and e-commerce, and is currently helping the World Bank design and implement an integrated loan administration platform.

**Jayanta Basak** *IBM Research Division, IBM India Research Laboratory, Block I, Indian Institute of Technology (IIT), Hauz Khas, New Delhi 110016, India (bjayanta@in.ibm.com).* Dr. Basak received a B.E. degree in electronics and telecommunication engineering from Jadavpur University, Calcutta; an M.E. degree in computer science and engineering from the Indian Institute of Science (IISc), Bangalore; and a Ph.D. degree from the Indian Statistical Institute (ISI), Calcutta, in 1987, 1989, and 1995, respectively. He served as a computer engineer in the Knowledge Based Computing Systems project of ISI from 1989 to 1992, and on the ISI faculty from 1992 to 2000. Dr. Basak was a visiting scientist at the Robotics Institute of Carnegie Mellon University from 1991 to 1992 and a researcher in the RIKEN Brain Science Institute, Japan, from 1997 to 1998. Since September 2000, he has been a Research Staff Member of the IBM India Research Laboratory, Delhi. He received Young Scientist awards from the Indian Science Congress Association (ISCA), the Indian National Science Academy (INSA), and the International Neural Networks Society (INNS) in 1994, 1996, and 2000, respectively. Dr. Basak's research interests include neural networks and soft computing, pattern recognition, vision, image processing, and e-commerce.

**Vivek Jain** *IBM Research Division, IBM India Research Laboratory, Block I, Indian Institute of Technology (IIT), Hauz Khas, New Delhi 110016, India (jvivek@in.ibm.com).* Mr. Jain received his B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, in 1994 and his postgraduate diploma in management from the Indian Institute of Management, Bangalore, in 1996. Thereafter, he worked with the ICICI Securities and Finance Company in the area of financial and business analysis. At IBM Research, his areas of interest since June 2000 have included statistical analysis, learning algorithms, and fuzzy logic, with a focus on e-commerce applications, extracting business intelligence from enterprise-wide data, and marketing research. He is currently leading the business development efforts for the IBM India Research Laboratory in the Indian and southeast Asian markets.

**Ravi Kothari** *IBM Research Division, IBM India Research Laboratory, Block I, Indian Institute of Technology (IIT), Hauz Khas, New Delhi 110016, India (rkothari@in.ibm.com).* Dr. Kothari received the B.E. degree (with distinction) from Birla Institute of Technology, India, the M.S. degree from Louisiana State University, and the Ph.D. degree from West Virginia University, all in electrical engineering. He joined the Department of Electrical and Computer Engineering (ECECS) and Computer Science of the University of Cincinnati (UC), Ohio, as an Assistant Professor; he later became a tenured Associate Professor and Director of the Artificial Neural Systems Laboratory. Since June of 2002, Dr. Kothari has worked in the IBM India Research Laboratory, New Delhi. His theoretical areas of interest include pattern recognition, machine learning, data mining, self-organization, and dimensionality reduction. He is currently applying these technologies in commerce, life sciences, and other domains. Dr. Kothari received the William E. Restemeyer Teaching Excellence Award from the Department of ECECS at UC in 1994 and the Eta Kappa Nu Outstanding Professor of the Year Award from the Department of ECECS at UC in 1995. He serves as an Associate Editor of the *IEEE Transactions on Knowledge and Data Engineering* and the *IEEE Transactions on Neural Networks*, as an Editorial Board member of the *Pattern Analysis and Applications Journal*, and as a member of the program committees of numerous conferences. Dr. Kothari is a member of Sigma Xi, Eta Kappa Nu, and Phi Kappa Phi; he is an IEEE Distinguished Visitor (2003–2005).

**Manoj Kumar** *IBM Corporate Division, Route 100, Somers, New York 10589 (manoj1@us.ibm.com).* Dr. Kumar is the Director of Corporate Technology Evaluation at IBM. He served as the Director of the IBM India Research Laboratory from 2000 to 2003. From 1983 to 2000, Dr. Kumar worked at the IBM Thomas J. Watson Research Center in Yorktown Heights, New York, where he led projects in electronic commerce, video delivery and encoding technologies, and architecture and design of processors and of parallel computers. Dr. Kumar received his B.Tech. degree from the Indian Institute of Technology, Kanpur, in 1979, and his M.S. and Ph.D. degrees from Rice University in 1981 and 1984, respectively, all in electrical engineering.

**Parul A. Mittal** *IBM Research Division, IBM India Research Laboratory, Block I, Indian Institute of Technology (IIT), Hauz Khas, New Delhi 110016, India (mparul@in.ibm.com).* Ms. Mittal is a Research Staff Member at the IBM India Research Laboratory, Delhi. She holds a B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, and a Master of Science degree from the University of Michigan in computer science. Prior to joining IBM, she worked at Hughes Software Systems in India. Her primary research interests are in the area of e-commerce, knowledge management, and e-governance.

**Natwar Modani** *IBM Research Division, IBM India Research Laboratory, Block I, Indian Institute of Technology (IIT), Hauz Khas, New Delhi 110016, India (namodani@in.ibm.com).* Mr. Modani received an M.E. degree from the Indian Institute of Science, with specialization in high-speed networks. At the IBM India Research Laboratory, his area of interest is e-commerce.

**K. Ravikumar** *(biography not available)*

**Yogish Sabharwal** *IBM Research Division, IBM India Research Laboratory, Block I, Indian Institute of Technology (IIT), Hauz Khas, New Delhi 110016, India (ysabharwal@in.ibm.com).* Mr. Sabharwal received his bachelor's degree in mathematics in 1994 and his master's degree in computer applications in 1997, both from Delhi University. He is currently pursuing a Ph.D. degree at the Indian Institute of Technology, Delhi. Mr. Sabharwal has worked with Hughes Software Systems and Nortel Networks in the areas of networking and telecommunications and is currently working on e-commerce technology at the IBM India Research Laboratory.

**R. Sureka** *(biography not available)*



Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.