Operations Research

Rating Customers According to Their Promptness to Adopt New Products

Dorit S. Hochbaum, Erick Moreno-Centeno, Phillip Yelland, Rodolfo A. Catena,

Rating Customers According to Their Promptness to Adopt New Products

Dorit S. Hochbaum
Department of Industrial Engineering and Operations Research, University of California, Berkeley, Berkeley, California 94720,
hochbaum@ieor.berkeley.edu

Erick Moreno-Centeno
Department of Industrial and Systems Engineering, Texas A&M University, College Station, Texas 77845, e.moreno@tamu.edu

Phillip Yelland
Google Inc., Mountain View, California 94043, phillip.yelland@gmail.com

Rodolfo A. Catena
The SPHERE Institute, Burlingame, California 94010, rodolfo.catena@gmail.com

Databases are a significant source of information in organizations and play a major role in managerial decision-making. This study considers how to process commercial data on customer purchasing timing to provide insights on the rate of new product adoption by the company's consumers. Specifically, we show how to use the separation-deviation model (SD-model) to rate customers according to their proclivity for adopting products for a given line of high-tech products. We provide a novel interpretation of the SD-model as a unidimensional scaling technique and show that, in this context, it outperforms several dimension-reduction and scaling techniques. We analyze the results with respect to various dimensions of the customer base and report on the generated insights.

1. Introduction

Databases are a significant source of information in organizations and play a major role in managerial decision making. From commercial data, organizations derive information about their customers and use it to hone their competitive strategies.

The rating of customers with respect to the promptness to adopt new products is a compelling exercise, because it allows companies to define appropriate actions for the launch of a new product into the marketplace. *Innovators*, customers that adopt technology promptly, are often the main target of a firm's marketing efforts of new products. Because the innovators tend to influence the remaining potential adopters, that is, the *majority*, firms tend to allocate more marketing efforts and resources toward the innovators than toward the majority (Mahajan and Muller 1998). Therefore, knowing the customers' *adoption promptness* allows companies to focus their marketing to innovators effectively. In addition, customer rating is the first step to be able to perform studies that link individual characteristics (such as age, gender, usage rate, and loyalty) to the adoption promptness.

Rating the customers' adoption promptness is particularly important in high tech markets, where products generally have short—and indeed shrinking—life cycles

(Talluri et al. 1998). For example, whereas memory semiconductor chips had a life of mature product lasting approximately five years in the early 90s, this had shrunk to one year in the early 2000s, (see, e.g., Figure 1 for product life cycles in the semiconductor industry).

The motivation of this paper is to solve the *customer rating problem*: Given data on a set of customers, a set of products on a given product line, and the purchase times of each customer-product pair, the customer rating problem is to rate each customer according to his/her adoption promptness. The proposed methodology is illustrated for commercial data from Sun Microsystems by rating the adoption promptness of some of their customers.

Our focus is on the customer rating problem where the information available is incomplete; that is, there are customers who do not purchase every product. This *incomplete information scenario* is a feature of the Sun database and it is typical for other companies, as well, that not all customers buy every product.

The customer rating problem is addressed here with Hochbaum's separation-deviation model (Hochbaum 2004), which has previously been used in contexts such as group decision making (Hochbaum and Levin 2006) and country-credit risk rating (Hochbaum and Moreno-Centeno 2008). This optimization model aims to minimize the sum of the

penalties on deviating from *priors* and from *pairwise comparisons*. In the customer rating context, the priors are the customers' product purchase timings, and the pairwise comparisons are derived from the relative difference in timing of customers' product purchases. The model is efficiently solvable, resulting in a scalar value for each customer representing their overall score of adoption promptness. Note that the adoption behavior of customers usually differs over different lines of products. Therefore, an important assumption of the model is the homogeneity of the studied products. That is, the model assumes that the products belong to the same product line, and thus the customers have similar adoption behavior on the studied products. (Even under this assumption, different products are likely to indicate different timings and pairwise comparisons.) Under this assumption, it is appropriate to give a single rating for each customer; this rating indicates the promptness to adopt a new product on this particular product line.

Two of the main contributions of this paper are (1) applying Hochbaum's separation-deviation model to the customer rating problem and (2) reinterpreting the separation-deviation model as a unidimensional scaling technique, thus presenting the model as an alternative to well-known dimension-reduction methodologies (in the special case where the data needs to be represented/summarized in only one dimension), including unidimensional scaling, principal component analysis, factor analysis, and averaging.

This paper is organized as follows. Section 2 reviews how the customer rating problem has been previously addressed in the literature and reviews several scaling and dimension-reduction methodologies that can be applied to solve the customer rating problem. Section 3 reviews the separation-deviation model (hereafter referred to as SD-model) and indicates how the SD-model differs from other techniques. Section 4 compares, in simulated scenarios, the performance of the SD-model to that of unidimensional scaling. Specifically, §4 compares the performance of the two methods in simulated scenarios where the correct adoption promptness of the customers is known in advance. Section 5 presents a study on commercial data from Sun Microsystems and reports the generated insights obtained by using our approach. Finally §6 gives some final remarks about the SD-model and its usefulness for other types of applications.

2. Literature Review

In general, the input to data-mining techniques consists of a collection of records that characterize customer purchase behavior, as well as other relevant customer characteristics such as age, gender, usage rate, loyalty status, etc. At an abstract level, many data-mining techniques attempt to explain customer behavior in terms of a meaningful subset of customer characteristics by identifying a function that maps a vector of customer attributes to a scalar value.

There are two main classes of data-mining techniques: those for *supervised learning* and those for *unsupervised learning*. The main objective of supervised learning techniques is to try to identify how to use independent variables (i.e., observable customer characteristics) to be able to predict an unobservable customer characteristic. These techniques require as input a customer database with preclassified customers. Some classical customer segmentation techniques that fall under this category are automatic interaction detector (AID) and its extensions (e.g., CHAID), linear regression and its generalizations (e.g., canonical analysis), discriminant analysis, conjoint analysis and its extensions (e.g., componential segmentation and *POSSE*—product optimization and selected segment evaluation), logistic regression, neural networks, etc. For the Sun Microsystems study we have no a priori labeling of the customers, i.e., we do not have a "training set." Therefore, the focus of this paper is on the unsupervised-learning problem of rating customers according to their adoption promptness.

In unsupervised-learning techniques, there is no preclassified set of customers. Thus, the unsupervised-learning techniques aim to determine the customer ratings from the unlabeled data. Unsupervised-learning techniques can be classified as cluster-analysis or dimension-reduction techniques.

Cluster-analysis techniques solve the following problem: Given a data set containing information about n objects, *cluster* these objects into groups, such that objects belonging to the same cluster are similar in some sense. Cluster-analysis methods such as K-means, hierarchical clustering, and Gaussian mixture models find a partition of the objects, so that the objects on each subset (cluster) share a common trait. We mention a clustering approach, based on maximum-cut clustering, for the customer segmentation problem (Rusmevichientong et al. 2004). Maximum-cut is an NP-hard problem, so the approach in Rusmevichientong et al. (2004) is to approximate maximum-cut with semidefinite programming. The output is not a full customer rating, but rather a classification of the customers only in early versus late adopters. We cannot compare directly clustering techniques to the scaling methodology proposed in this paper because the outputs are different. In particular, clustering techniques output a partition of the customers into clusters, whereas our methodology assigns a rating to each customer. Therefore we decided to compare our methodology with dimension-reduction techniques, whose output is straightforwardly comparable to that of our methodology.

Dimension-reduction techniques (DRTs) solve the following problem: Given an $n \times k$ matrix, R, find the $n \times k'$ matrix with $k' < k$ that best captures the content in the original matrix, according to a certain criterion. In the customer segmentation problem, R is the matrix containing the purchase times of k products by n customers, and the output is an $n \times 1$ vector that captures the relative "purchase ordering" of the customers. Some of the most widely used

DRTs are principal component analysis (PCA), factor analysis (FA), multidimensional scaling (MDS), and averaging. We give (below) a brief description of each technique. For an in-depth discussion of PCA, FA, and MDS, we refer the reader to Jollife (1986), Rummel (1970), and Torgerson (1952), respectively.

In essence, PCA seeks to reduce the dimension of the data by finding a few orthogonal linear combinations (the principal components) of the original variables with the largest variance. The first principal component is the linear combination with the largest variance; in this sense, it is the one-dimensional vector that best captures the information contained in the original data.

Factor analysis assumes that the measured variables depend on some unknown, and often unmeasurable, common factors. The goal of FA is to uncover such relations. Typical examples include variables defined as various test scores of individuals, because such scores are thought to be related to a common "intelligence" factor. Here the measured variables are the purchase times of a customer, and the unmeasurable factor of interest is the customer's proclivity for early adoption.

Given n items in a k-dimensional space and an $n \times n$ matrix of distances among the items, MDS produces a k'-dimensional, $k' < k$, representation of the items such that the pairwise distances among the n points in the new space are similar to the distances in the original data.

The customer rating problem can also be addressed by the (naive) DRT of the *averaging method*. Given a set of purchase times for customer i of product k, r_i^k, the rating obtained by the averaging method is given by $x_i^{\mathrm{avg}} = (\sum_{k \in R_i} r_i^k)/(|R_i|)$, where R_i is the set of products purchased by the ith customer.

For most DRTs, including PCA and FA, missing data pose serious problems (see Kosobud 1963, Afifi and Elashoff 1966, for example). In the customer rating problem, assuming full data is equivalent to assuming that all customers purchased every product. As discussed in §1, this does not hold in general, and in particular it does not hold for Sun's data. These DRTs, PCA, and FA require that the missing values are estimated and artificially imputed. (In statistics, imputation is the substitution of some value for a missing data point or a missing component of a data point.) In contrast, modern versions of MDS (thoroughly reviewed below) are designed specifically to handle missing data without the need of imputation. Although modern versions of PCA and FA versions do, in some sense, work on imputed values as well, they require that the imputed values are consistent with an underlying stochastic model for the data. In the problem herein considered, there is not enough data to fit an underlying stochastic model. Thus, to be able to use PCA and FA to solve our problem, we imputed the missing values using a simple nearest-neighbor missing-data-recovery method. Although this imputation method has been shown to be appropriate in the treatment of missing data (Huang and Zhu 2002, Hruschka et al. 2003), this

was not the case in our study. Specifically, when using PCA and FA on the imputed data to solve our problem, their performances were dominated by those of the SD-model and MDS. (We note that there might be other imputation methods that could potentially lead to better results.) Therefore, we will only present (in §4) the comparison of the performances of SD-model to MDS. Because the performance of the average method was also dominated by that of the SD-model and MDS, we decided also not to include such a performance comparison.

2.1. Review of Multidimensional Scaling

Multidimensional scaling (MDS) is a set of related techniques used for representing the similarities and dissimilarities among pairs of objects as distances between points on a low-dimensional space. MDS models aim to approximate given nonnegative dissimilarities, δ_{ij}, among pairs of objects, i, j, by distances between points in an m-dimensional MDS configuration X. Here X, the *configuration*, is an $n \times m$ matrix with the coordinates of the n objects in $\Re^m$. Most MDS techniques assume that the dissimilarity matrix $[\delta_{ij}]$ is symmetric; we review two important exceptions below. The most common function to measure the fit between the given dissimilarities, δ_{ij}, and distances, $d_{ij}(X)$, is STRESS, defined by

$$\mathrm{STRESS}(X) \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (\delta_{ij} - d_{ij}(X))^2, \qquad (1)$$

where w_{ij} is a given nonnegative weight reflecting the importance or precision of the dissimilarity δ_{ij}. Note that w_{ij} can be set to 0 if δ_{ij} is unknown. $d_{ij}(X)$ is a vector norm, defined as

$$d_{ij}(X) = \left[\sum_{s=1}^{m} |x_{is} - x_{js}|^q \right]^{1/q}$$

with given parameter $q \geqslant 1$. Usually, $d_{ij}(X)$ is the L_2 norm ($q = 2$) or the L_1 norm ($q = 1$).

Finding a global minimum of (1) is a hard optimization problem because STRESS is a nonlinear nonconvex function with respect to X, and thus optimization algorithms can converge to local minima (see, for example, de Leeuw 1977, Groenen et al. 1999, Alexander et al. 2005).

In a useful MDS technique, the *three-way MDS*, for each pair of objects we are given K dissimilarity measures from different "replications" (e.g., repeated measures, different experimental conditions, multiple raters, etc.). This technique is referred to as *three*-way MDS because the input is a three-dimensional matrix δ_{ij}^k, as opposed to the two-dimensional matrix in "classic" MDS. The objective function of three-way MDS is defined as (de Leeuw 1977),

$$\mathrm{3WAY\text{-}STRESS}(X) \equiv \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^k (\delta_{ij}^k - d_{ij}(X))^2. \qquad (2)$$

Unidimensional scaling (UDS) is the important one-dimensional case of MDS where the configuration X is an $n \times 1$ matrix. Therefore, UDS seeks to approximate the given dissimilarities by distances between points in a one-dimensional space. Unidimensional scaling has been used successfully in several contexts (see, for example, Fisher 1922, Robinson 1951, Ge et al. 2005). Unidimensional scaling has been studied mainly as a model for object sequencing and seriation (Hubert et al. 2001, Brusco and Stahl 2005b), thus its relevance to the problem concerning this paper. Unidimensional scaling is a hard optimization problem, and combinatorial techniques (e.g., branch and bound and dynamic programming) are only able to optimally solve instances of up to 30 objects; see, for example, Lau et al. (1998), Brusco (2002), Hubert et al. (2002), Brusco and Stahl (2005c).

In our particular application, rating customers according to their adoption promptness, the input data is a matrix R with r_i^k giving the adoption time (relative to product launch) of customer i for product k. This matrix is, in general, incomplete and has many missing elements. The objective is to assign each customer i to a scale $\mathbf{x}$ such that x_i most accurately recovers the across-customer ordering of product adoption times within any product. To solve our problem, we can set up the following three-way UDS problem:

$$\min_{\mathbf{x}} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^k (|r_i^k - r_j^k| - |x_i - x_j|)^2. \tag{3}$$

Here the interpretation is that product k gives a pairwise dissimilarity, $|r_i^k - r_j^k|$, among a pair of customers i and j the purchased product k. Then, the objective is that customers with low (high) dissimilarities have similar (dissimilar) adoption promptness and should be placed "close (far) to each other" in the desired scale $\mathbf{x}$.

We note a couple of drawbacks of formulating our customer rating problem as the three-way UDS problem (3), and later introduce our scaling methodology which, addresses these drawbacks.

1. As mentioned earlier, finding the optimal solution to (3) is a hard optimization problem because the objective is nonconvex (Groenen et al. 1999); current optimization techniques are only able to optimally solve instances of at most 30 objects.

2. By calculating the dissimilarities as $|r_i^k - r_j^k|$, problem (3) ignores the so-called *directionality of dominance*, that is, the sign of $(r_i^k - r_j^k)$. In particular, problem (3) does not capture the information regarding which customer adopted product k earlier. Note that this information is very relevant in the customer rating problem.

A closely related observation is that, given an optimal solution, $\mathbf{x}^*$, to (3), $-\mathbf{x}^*$ is also an optimal solution to (3). Thus, by solving (3), we get a rating of the customers, but we do not know whether a higher rating means a greater adoption promptness or vice versa.

Although the vast majority of the papers in the UDS literature assume that the given dissimilarities are nonnegative

and symmetric, there are two papers (Hubert et al. 2001 and Brusco and Stahl 2005a) that consider the case where the dissimilarities are given in a complete skew-symmetric matrix (i.e., $\delta_{ij} = -\delta_{ji}$).

Because these approaches consider only one matrix $[\delta_{ij}]$ and this matrix is complete, these are not applicable to the customer rating problem. Indeed, the approach presented in this paper is a nice generalization of one of these approaches. We briefly discuss the approaches presented in Hubert et al. (2001) and Brusco and Stahl (2005a) and refer to the original papers for further details.

Hubert et al. (2001) observe that a skew-symmetric matrix contains two distinct types of information between any pair of objects: degree of dissimilarity, $|\delta_{ij}|$, and directionality of dominance, $\text{sign}(\delta_{ij})$. They consider two approaches to sequencing the objects. The first approach consists of finding the object ordering π such that the matrix $[\delta_{\pi(i)\pi(j)}]$ has the maximum sum of above-diagonal entries. Hubert et al. note that this problem is exactly the minimum feedback arc set problem, which is NP-hard. The second approach proposed in Hubert et al. (2001) is to solve the following problem,

$$\min_{\mathbf{x}} \sum_{i=1}^{n} \sum_{j=1}^{n} (\delta_{ij} - (x_i - x_j))^2, \tag{4}$$

where the dissimilarity matrix $[\delta_{ij}]$ is assumed to be skew symmetric and has no missing entries. Hubert et al. give an analytic solution to problem (4); Hochbaum and Moreno-Centeno (2008) give a generalization of this result to the case of multiple dissimilarity matrices (but still no missing entries).

Brusco and Stahl (2005a) also differentiate between the degree of dissimilarity, $|\delta_{ij}|$, and directionality of dominance, $\text{sign}(\delta_{ij})$. They propose a bicriteria optimization problem that balances between these two types of information. Although interesting, this approach is not practical because the proposed solution technique is only able to determine the nondominated solutions for matrices up to size 20×20 (and can take as input only one skew-symmetric matrix).

3. The Separation-Deviation Model

3.1. Review of the Separation-Deviation Model

The SD-model was proposed by Hochbaum (2004, 2006). The inputs for the separation-deviation model are a set of objects $\{1, \ldots, n\}$, for each object a set of prior ratings r_i^k for $k = 1 \ldots K$, and a set of pairwise comparisons δ_{ij}^k for $k = 1 \ldots K$ for each pair of objects. These pairwise comparisons are skew-symmetric, that is $\delta_{ij}^k = -\delta_{ji}^k$. The SD-model aims to assign each object a rating $\mathbf{x}_i$, such that $\mathbf{x}_i$ is as close as possible to the given prior ratings and the difference in the ratings of each pair of objects is as close as possible to the given pairwise comparisons.

Let the variable x_i be the rating of the ith object, and the variable z_{ij} be the difference x_i and x_j. The convex optimization formulation of the SD-model is

$$\min_{\mathbf{x}, \mathbf{z}} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^k f_{ij}^k (z_{ij} - \delta_{ij}^k) + \sum_{k=1}^{K} \sum_{i=1}^{n} v_i^k g_i^k (x_i - r_i^k) \quad (5a)$$

$$\text{s.t. } z_{ij} = x_i - x_j \quad (i = 1, \ldots, n; \ j = i+1, \ldots, n). \quad (5b)$$

The penalty function $f_{ij}^k()$ for disagreeing from the kth pairwise comparison between the ith and jth objects is a convex function of $z_{ij} - \delta_{ij}^k$. The total sum of these penalties, $\sum_k \sum_i \sum_j f_{ij}^k (z_{ij} - \delta_{ij}^k)$ is called the *separation penalty*. As in MDS, w_{ij}^k are given nonnegative weights reflecting the importance or precision of δ_{ij}^k and are set to 0 if δ_{ij}^k is unknown.

The penalty function $g_i^k()$ for disagreeing from the kth prior rating on the ith object is a convex function of $x_i - r_i^k$. The total sum of these penalties $\sum_k \sum_i v_i^k g_i^k (x_i - r_i^k)$ is called the *deviation penalty*. The v_i^k are given nonnegative weights reflecting the importance or precision of the purchase time r_i^k and are set to 0 if the ith customer did not buy the kth product.

It was proved in Hochbaum and Levin (2006) that problem (5) is a special case of the convex dual of the minimum cost network flow (CDMCNF) problem. As such it is solvable by the efficient polynomial-time algorithm devised in Ahuja et al. (2003).

3.2. Comparison Between UDS and the SD-Model

A simplified version of the SD-model is to solve problem (6). In this subsection, we present this simplified version to allow for a quick comparison with MDS/UDS.

$$\min_{\mathbf{x}} \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^k f_{ij}^k (\delta_{ij}^k - (x_i - x_j)), \quad (6)$$

where, for the customer rating problem, $\delta_{ij}^k \equiv r_i^k - r_j^k$, and thus, for each product k, $[\delta_{ij}^k]$ is a (possibly incomplete) skew-symmetric matrix. w_{ij}^k is a nonnegative weight reflecting the importance or precision of the dissimilarity δ_{ij}^k (w_{ij}^k is set to 0 if δ_{ij}^k is unknown). Also, each $f_{ij}^k(\cdot)$ is a given of convex function.

In MDS terminology, problem (6) is a three-way unidimensional scaling problem where the K dissimilarity matrices are skew symmetric. In contrast to all of the MDS literature, where $f_{ij}^k(\cdot)$ are either the quadratic function or the absolute value function, in the separation-deviation model these functions can differ from each other and may be any convex function. In contrast to problem (3)—the direct application of UDS to the customer rating problem—problem (6) is solvable in polynomial time and does not ignore the directionality of the dominance. In contrast to the approaches in Hubert et al. (2001) and Brusco and Stahl (2005a) for skew-symmetric matrices, problem (6) accepts multiple and incomplete dissimilarity matrices and is solvable in polynomial time.

3.3. Customer Rating via the Separation-Deviation Model

Consider a population of customers, identified by the index i, who may elect to purchase products (on a given product line) indexed by $k \in \{1, \ldots, K\}$ over a period comprising a number of periods (months). Let r_i^k be the first month (if any) in which customer i purchased product k. Each of the n customers is associated with a K-dimensional vector $\mathbf{r}_i = (r_i^1, \ldots, r_i^K)$, recording the first month in which he or she bought the different products. In the event that the customer did not purchase a product, the corresponding entry in the vector is regarded as "missing." The model appropriately (and seamlessly to the user) deals with this missing information.

As shown in Figure 1, the life cycles of products tend to shrink over time. This is particularly the case in the high-tech industry. When the products compared span a few years, one might want to mitigate the extra weight that earlier products have due to their availability for purchase over a larger period of time. This is done by calibrating the values of r_i^k; that is, the input to the SD-model should be given in terms of the relative time position within the span of the k product's life cycle, instead of the absolute values in term of months. Indeed, throughout this paper we assume that the life cycles of the K products are of equal length. This is without loss of generality, because if the products have different life cycle lengths these are calibrated to equal value (say $[0, 1]$) by dividing the month of purchase by the cycle length, in months, of the respective product.

One of the important features of the separation-deviation model is that the model takes as input a collection of pairwise comparisons between the objects (customers) to be classified. That is, a single customer-pair can have several, possibly conflicting, pairwise comparisons. In this particular application, the SD-model uses the purchase times to create pairwise comparisons among the different customers.

Figure 1. Shrinking product life cycles in the semiconductor industry over time.
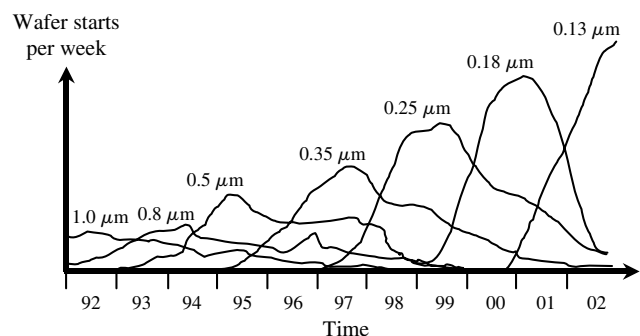
EXAMPLE 1. Consider the input given in the following table:

r_i^k	a	b	c
A	8	—	1
B	—	2	6
C	9	1	5

The first row of the table says that customer A bought products a and c eight months and one month after each was launched; customer A did not buy product b. In addition to considering the information given in the above table, the model explicitly generates five pairwise comparisons that represent the difference in the adoption promptness between each pair of customers. For instance, the model will explicitly use the fact that customer A bought product a one month before customer C.

The main motivations for considering pairwise comparisons are the following:

1. We are interested in differentiating between customers that buy earlier and customers that buy later; this is a question about the relative purchase times between customers. In this respect it is important to emphasize that we are not concerned with the problem of predicting the time when a certain customer will buy a given product (the "absolute" purchase times of each customer).

2. Although the specific months of purchase for different products might have a high variation, the relative difference in the purchase times between two given customers might have less variation. So, for example, say that Alice buys two products in months 1 and 100, respectively, and Bob buys the same products in months 3 and 110, respectively. Just looking at Alice's purchases, it would be unclear to determine if she is an early adopter or a late adopter; however, when considering that she always bought the products earlier than Bob, we can be certain that she adopts new products faster than Bob.

In the application described in this paper, pairwise comparisons between customers are derived from the observed first purchase times described above. Specifically, let r_i^k and r_j^k be the observed first purchase times of customers i and j, respectively, both of whom bought product k. Then the pairwise comparison between the two customers is defined as $\delta_{ij}^k = r_i^k - r_j^k$; note that comparisons are skew symmetric, because $\delta_{ij}^k = -\delta_{ji}^k$. Ultimately, the output of the SD-model is a set of values, or *ratings*, associated with each object so that the difference between each pair deviates as little as possible from the input pairwise comparisons.

In this study, we use for the penalty functions $f_{ij}^k()$ and $g_i^k()$ the absolute value functions (problem (7) below) and the quadratic convex penalty functions (problem (8) below). We set w_{ij}^k equal to 1 if both customers i and j bought product k, and set w_{ij}^k equal to 0 otherwise. Similarly, we

set v_i^k equal to 1 if customers i bought product k, and set v_i^k equal to 0 otherwise.

$$\min_{\mathbf{x}, \mathbf{z}} \ M \cdot \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^k |z_{ij} - \delta_{ij}^k| + \sum_{k=1}^{K}\sum_{i=1}^{n} v_i^k |x_i - r_i^k| \quad (7a)$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n; \ j = i+1, \ldots, n). \quad (7b)$$

$$\min_{\mathbf{x}, \mathbf{z}} \ M \cdot \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^k (z_{ij} - \delta_{ij}^k)^2 + \sum_{k=1}^{K}\sum_{i=1}^{n} v_i^k (x_i - r_i^k)^2 \quad (8a)$$

$$\text{s.t.} \quad z_{ij} = x_i - x_j \quad (i = 1, \ldots, n; \ j = i+1, \ldots, n). \quad (8b)$$

In problems (7) and (8) the parameter M is chosen so that the separation penalty is *lexicographically* more important than the deviation penalty. By lexicographically more important, we mean that the separation penalty is the dominant term in the optimization problem so that the deviation penalty is only used to choose among the feasible solutions with minimum separation penalty. We set the separation penalty to be lexicographically more important than the deviation penalty because it better represents our objective to determine the relative purchase-time ordering of the customers (as opposed to predicting the absolute purchase times of the customers).

A property of the separation deviation model demonstrated in Hochbaum and Moreno-Centeno (2008) shows that when $\{r_i^k\}$ has no missing entries (in the customer rating context this means that all products are purchased by all customers), using the absolute value penalty function guarantees that the output rating will agree with the rating implied by the majority of the raters (here the raters are the products). We denote the optimal solution to problems (7) and (8) as $\mathbf{x}^{|SD|}$ and $\mathbf{x}^{SD^2}$, respectively.

EXAMPLE 2. The following table provides the optimal solutions of problems (7) and (8) for the data in Example 1. The column $\mathbf{x}^{avg}$ gives the row average of the inputs.

	Inputs			Outputs		
	a	b	c	$\mathbf{x}^{|SD|}$	$\mathbf{x}^{SD^2}$	$\mathbf{x}^{avg}$
A	8	—	1	1	2.9	4.5
B	—	2	6	6	5.6	4.0
C	9	1	5	5	5.0	5.0

As the above table shows, both $\mathbf{x}^{|SD|}$ and $\mathbf{x}^{SD^2}$ preserve the order of the customers implied by all the products purchased. On the other hand, taking the row average does not preserve such order; indeed, it contradicts the ordering of customers (B, C) according to product b, and the ordering of the customers (A, B) and (B, C) according to product c.

Below we make some relevant observations regarding the SD-model.

• The SD-model is an approach for unsupervised learning, and as such it does not require a training set.

- A particular advantage of the SD-model is that in addition to working well in situations with complete information (see Hochbaum and Moreno-Centeno 2008), it works well (without any need for data preprocessing) in situations where we have incomplete information. This is particularly prominent in the application studied here, where the information matrix is sparse. Indeed, in the Sun Microsystems database, only 165 of 1,916 customers bought all the products considered in this study, and 1,132 customers only bought one product.

- The SD-model can take subjective, or less than entirely reliable, judgments as input and calibrate those inputs with appropriate confidence levels.

- The SD-model is solvable in polynomial time for any convex penalty functions $f_{ij}^k()$ and $g_i^k()$.

- The segmentation achieved by the SD-model is the "best" according to a defined metric.

- The SD-model makes obvious the discrepancies that exist between the inputs. Outliers are made explicit and may be used to improve insights into customer behavior. In particular, by identifying the highest penalty terms one can detect outliers in both the pairwise comparisons and in the prior ratings.

- The SD-model does not rely on specified distributions for different classes, and there is no requirement of any specific sample size.

4. Performance Assessment on Simulated Scenarios

This section assesses the performance of the SD-model under several different simulated scenarios and compares its performance to that of three-way UDS (problem (3)).

We denote as $\mathbf{x}^{\text{UDS}}$ the customer rating obtained using three-way UDS, that is, the solution to problem (3). Recall that obtaining the optimal solution to problem (3) is only possible (with current optimization techniques) for $n \leqslant 30$ (Hubert et al. 2001). Although specialized heuristics to solve UDS are available (Hubert et al. 2002 survey these heuristics), none of them apply to the three-way UDS or to the weighted UDS (that is, all of the heuristics assume that the data is complete and $w_{ij} = 1$ for $i, j = 1, \ldots, n$). Therefore to find a heuristic solution to problem (3), we used Matlab's heuristic to solve the weighted MDS problem. Strictly speaking, Matlab's heuristic was designed to minimize problem (1) (that is, it only accepts one dissimilarity matrix). However, as shown in de Leeuw (1977), when using the quadratic function as penalty function, minimizing (2) can be reduced to the problem of minimizing (1).

Each scenario represents a different customer purchase-timing behavior and consists of 600 customers each buying up to four products. We associate a different purchase-time distribution with each customer-product pair. By letting the purchase-time distributions depend on both the customer and the product, we are able to simulate scenarios where the products have different life cycles and characteristics.

In these scenarios, the customers' purchase times may have different expected values and/or variances depending on the product under consideration.

We simulated the purchase time of each product by each customer using the gamma distribution, which is commonly used to simulate "customer arrival times." Let c and p represent the index of the customer and product, respectively. We used seven different expected purchase times $\{c, c + 2p, c + 5p, c + 50p, cp, 10cp, 50cp\}$, and 11 different variances $\{10, 50, 5c, 10c, 50c, 5p, 10p, 50p, 5c + 5p, 10c + 10p, 50c + 50p\}$. Overall, we simulated 77 difference scenarios, one for each possible mean-variance combination. For example, in the scenario having cp mean and $5p + 5c$ variance, the purchase time of the jth product by the ith customer had an expected value of ij and a variance of $5i + 5j$. Note that in all of these scenarios, customers with lower indices adopt new products earlier. That is, given two customers i, j such that $i < j$, then, for any given product, customer i has an earlier expected purchase time than customer j. Thus, for every one of the 77 scenarios, the customers are ordered with respect to their adoption promptness. In particular, the *true ranking*, x_i^{T}, of the ith customer is equal to i. Note that the simulated customers behave similarly across products in that lower-index customers adopt (in expected value) earlier than higher-index customers for any given product; this is consistent with the assumption that all the products belong to the same product line.

To measure how successful the SD-model is in recovering the true ranking vector, $\mathbf{x}^{\text{T}}$, of the customers, we used Kendall's *Tau* rank-correlation coefficient. This coefficient provides a measure of the degree of correspondence between two vectors. In particular, it assesses how well the order (i.e., rank) of the elements of the vectors is preserved. In Appendix A we provide a description of Kendall's *Tau* rank-correlation coefficient. We note that, as an alternative to three-way UDS (problem (3) and the SD model (problems (7) and (8)), we could instead find the customer rating vector that maximizes Kendall's *Tau* rank-correlation coefficient. We decided not to do so because (1) finding such a vector is NP-hard (Bartholdi et al. 1989) and (2) this objective would ignore the degree of dissimilarity between the adoption times of the customers. On the other hand, we believe that Kendall's *Tau* rank-correlation coefficient is appropriate to measure how well the customer ratings recovered x_i^{T}. (Notice that x_i^{T} gives the true ordering of the customers and does not give a degree of dissimilarity between the customers.)

Recall that this paper focuses on the case where the data available is incomplete; that is, there are customers who did not purchase every product. To generate incomplete data, we first simulated the complete data; that is, we simulated the purchase times of every customer-product pair and then deleted some of the purchase times at random. In Sun's data 59%, 23%, 9%, and 9% of the customers bought one, two, three, and four products, respectively. We mimicked

Table 1.   Average *Tau* correlation coefficients between $\mathbf{x}^T$ and $\mathbf{x}^{|SD|}$.

τ	10	50	5c	10c	50c	5p	10p	50p	5c + 5p	10c + 10p	50c + 50p
c	0.9913	0.9784	0.8884	0.8434	0.6934	0.9902	0.9857	0.9662	0.8873	0.8433	0.6903
c + 2p	0.9911	0.9783	0.8872	0.8445	0.6947	0.9903	0.9855	0.9666	0.8879	0.8450	0.6903
c + 5p	0.9912	0.9785	0.8886	0.8455	0.6900	0.9905	0.9856	0.9666	0.8851	0.8444	0.6906
c + 50p	0.9913	0.9784	0.8877	0.8448	0.6847	0.9904	0.9856	0.9665	0.8863	0.8430	0.6830
cp	0.8427	0.8386	0.8271	0.8198	0.7745	0.8395	0.8416	0.8428	0.8270	0.8161	0.7766
10cp	0.8431	0.8449	0.8432	0.8419	0.8403	0.8439	0.8424	0.8441	0.8438	0.8426	0.8391
50cp	0.8426	0.8443	0.8433	0.8437	0.8452	0.8437	0.8433	0.8438	0.8451	0.8462	0.8423

this data by deleting the entries with this empirical distribution. In particular, each customer had a probability of 0.59, 0.23, 0.09, and 0.09 of buying one, two, three, and four products. For each customer the purchased products were chosen uniformly at random.

To summarize, the performance assessment of the SD-model on each of the 77 scenarios was executed as follows:

Step 1: Repeat 30 times:

Step 1.1: Simulate the purchase-time data of four products by 600 customers.

Step 1.2: Delete some of the purchase times at random to obtain incomplete information.

Step 1.3: Solve for $\mathbf{x}^{|SD|}$, $\mathbf{x}^{SD^2}$, and $\mathbf{x}^{UDS}$.

Step 1.4: Compute the Tau correlation coefficient between $\mathbf{x}^T$ (the true customer ranking) and each of $\mathbf{x}^{|SD|}$, $\mathbf{x}^{SD^2}$, and $\mathbf{x}^{UDS}$.

Step 2: Calculate the average and standard deviation of the 30 Tau correlation coefficients (with $\mathbf{x}^T$) computed for each of $\mathbf{x}^{|SD|}$, $\mathbf{x}^{SD^2}$, and $\mathbf{x}^{UDS}$.

Tables 1–3 give, for each of the 77 scenarios, the average Tau correlation coefficient between $\mathbf{x}^T$ and $\mathbf{x}^{|SD|}$, $\mathbf{x}^{SD^2}$, and $\mathbf{x}^{UDS}$, respectively.

To compare the performances of these methods, Tables 4–6 provide the average differences between the

correlation coefficients achieved by the different methods. For example, each entry in Table 4 is the difference between the corresponding entries of Tables 1 and 3. In Tables 4–6, the numbers given in **bold** are those that are at least three standard deviations above (or below) zero. Therefore the scenarios with **bold** entry are those where one method significantly outperforms the other; whereas, in the rest of the scenarios, the performance of both methods is essentially the same.

The results in Tables 4 and 5 provide evidence that the SD-model, irrespective of the penalty functions used, performs better than UDS on most of the 77 scenarios. In particular, the SD-model outperforms UDS in 46 out of the 77 scenarios; and in 36 of these 46 scenarios, the SD-model significantly outperforms UDS.

We now compare the performance of the SD-model using different penalty functions. For this purpose, Table 6 reports the average difference between the correlation coefficients (with respect to $\mathbf{x}^T$) for $\mathbf{x}^{|SD|}$ and $\mathbf{x}^{SD^2}$. We note that using the absolute-value penalty functions is only slightly better than using the quadratic-value penalty functions.

The simulation results indicate that the SD-model determines with high accuracy the true ranking of the customers with respect to their adoption promptness.

Table 2.   Average *Tau* correlation coefficients between $\mathbf{x}^T$ and $\mathbf{x}^{SD^2}$.

τ	10	50	5c	10c	50c	5p	10p	50p	5c + 5p	10c + 10p	50c + 50p
c	0.9915	0.9788	0.8903	0.8464	0.6958	0.9902	0.9858	0.9660	0.8891	0.8453	0.6928
c + 2p	0.9913	0.9787	0.8891	0.8468	0.6971	0.9902	0.9856	0.9666	0.8897	0.8474	0.6936
c + 5p	0.9914	0.9788	0.8904	0.8478	0.6925	0.9905	0.9856	0.9666	0.8873	0.8467	0.6932
c + 50p	0.9915	0.9789	0.8902	0.8474	0.6892	0.9904	0.9857	0.9665	0.8882	0.8452	0.6871
cp	0.8356	0.8322	0.8241	0.8169	0.7745	0.8333	0.8351	0.8365	0.8238	0.8140	0.7758
10cp	0.8349	0.8380	0.8357	0.8350	0.8339	0.8360	0.8348	0.8371	0.8368	0.8360	0.8331
50cp	0.8348	0.8373	0.8351	0.8366	0.8385	0.8362	0.8365	0.8368	0.8372	0.8383	0.8347

Table 3.   Average *Tau* correlation coefficients between $\mathbf{x}^T$ and $\mathbf{x}^{UDS}$.

τ	10	50	5c	10c	50c	5p	10p	50p	5c + 5p	10c + 10p	50c+50p
c	0.9787	0.9781	0.9766	0.9767	0.7198	0.7618	0.7613	0.9708	0.9712	0.9710	0.9704
c + 2p	0.7602	0.7697	0.7662	0.8871	0.8868	0.8879	0.8867	0.7610	0.7691	0.7645	0.8442
c + 5p	0.8450	0.8450	0.8458	0.7583	0.7677	0.7662	0.6893	0.6917	0.6769	0.6842	0.7287
c + 50p	0.7659	0.7678	0.9770	0.9782	0.9764	0.9756	0.7644	0.7664	0.7688	0.9764	0.9742
cp	0.9739	0.9740	0.7624	0.7676	0.7668	0.9603	0.9599	0.9617	0.9612	0.7665	0.7681
10cp	0.7674	0.8862	0.8874	0.8843	0.8854	0.7578	0.7150	0.7650	0.8443	0.8456	0.8436
50cp	0.8433	0.7188	0.7171	0.7195	0.6856	0.6887	0.6886	0.6816	0.7305	0.7628	0.7651

Table 4. Average difference between the correlation coefficients obtained by $\mathbf{x}^{|SD|}$ and $\mathbf{x}^{UDS}$.

τ	10	50	5c	10c	50c	5p	10p	50p	5c + 5p	10c + 10p	50c + 50p
c	0.0126	0.0003	−0.0882	−0.1333	−0.0265	**0.2283**	**0.2244**	−0.0046	−0.0840	−0.1277	−0.2801
c + 2p	**0.2310**	**0.2087**	**0.1210**	−0.0427	−0.1921	**0.1024**	**0.0988**	**0.2056**	**0.1188**	**0.0805**	−0.1539
c + 5p	**0.1462**	**0.1335**	**0.0427**	**0.0872**	−0.0778	**0.2243**	**0.2962**	**0.2748**	**0.2082**	**0.1603**	−0.0381
c + 50p	**0.2254**	**0.2106**	−0.0893	−0.1333	−0.2918	0.0148	**0.2212**	**0.2000**	**0.1175**	−0.1334	−0.2912
cp	−0.1312	−0.1354	0.0647	0.0522	0.0076	−0.1207	−0.1183	−0.1189	−0.1343	0.0496	0.0085
10cp	**0.0756**	−0.0413	−0.0442	−0.0424	−0.0451	0.0861	0.1274	**0.0792**	−0.0005	−0.0030	−0.0045
50cp	−0.0007	0.1255	0.1262	0.1241	**0.1596**	0.1549	0.1547	0.1622	0.1145	0.0834	**0.0773**

Note. Positive numbers indicate that $\mathbf{x}^{|SD|}$ has higher average correlation with $\mathbf{x}^{T}$ than $\mathbf{x}^{UDS}$.

Table 5. Average difference between the correlation coefficients obtained by $\mathbf{x}^{SD^2}$ and $\mathbf{x}^{UDS}$.

τ	10	50	5c	10c	50c	5p	10p	50p	5c + 5p	10c + 10p	50c + 50p
c	0.0128	0.0007	−0.0863	−0.1303	−0.0241	**0.2283**	**0.2244**	−0.0048	−0.0822	−0.1257	−0.2776
c + 2p	**0.2311**	**0.2090**	**0.1230**	−0.0403	−0.1896	**0.1024**	**0.0988**	**0.2056**	**0.1205**	**0.0829**	−0.1506
c + 5p	**0.1464**	**0.1338**	**0.0445**	**0.0894**	−0.0752	**0.2243**	**0.2963**	**0.2749**	**0.2103**	**0.1626**	−0.0356
c + 50p	**0.2256**	**0.2110**	−0.0869	−0.1307	−0.2872	0.0148	**0.2213**	**0.2000**	**0.1195**	−0.1312	−0.2871
cp	−0.1383	−0.1417	0.0617	0.0493	0.0077	−0.1270	−0.1248	−0.1251	−0.1375	0.0475	0.0077
10cp	**0.0675**	−0.0483	−0.0517	−0.0494	−0.0516	0.0782	0.1199	**0.0721**	−0.0075	−0.0096	−0.0104
50cp	−0.0085	0.1185	0.1179	0.1171	**0.1529**	0.1475	0.1479	0.1551	0.1067	**0.0754**	**0.0696**

Note. Positive numbers indicate that $\mathbf{x}^{SD^2}$ has higher average correlation with $\mathbf{x}^{T}$ than $\mathbf{x}^{UDS}$.

Table 6. Average difference between the correlation coefficients obtained by $\mathbf{x}^{|SD|}$ and $\mathbf{x}^{SD^2}$.

τ	10	50	5c	10c	50c	5p	10p	50p	5c + 5p	10c + 10p	50c + 50p
c	−0.0002	−0.0004	−0.0019	−0.0030	−0.0024	0.0000	−0.0001	0.0002	−0.0018	−0.0020	−0.0025
c + 2p	−0.0001	−0.0004	−0.0020	−0.0023	−0.0024	0.0000	0.0000	0.0000	−0.0017	−0.0024	−0.0033
c + 5p	−0.0002	−0.0003	−0.0018	−0.0022	−0.0026	0.0000	−0.0001	0.0000	−0.0021	−0.0023	−0.0026
c + 50p	−0.0002	−0.0004	−0.0024	−0.0026	−0.0045	0.0000	0.0000	0.0000	−0.0020	−0.0023	−0.0041
cp	0.0071	0.0064	0.0030	0.0029	0.0000	0.0063	0.0065	0.0062	0.0032	0.0021	0.0008
10cp	0.0081	0.0069	0.0075	0.0069	0.0064	**0.0079**	0.0075	0.0070	0.0070	0.0066	0.0059
50cp	0.0078	0.0070	**0.0082**	0.0071	0.0067	0.0074	0.0068	0.0071	**0.0079**	**0.0079**	0.0077

Note. Positive numbers indicate that $\mathbf{x}^{|SD|}$ has higher average correlation with $\mathbf{x}^{T}$ than $\mathbf{x}^{SD^2}$.

5. Rating Sun's Customers According to Their Adoption Promptness

5.1. Sun's Data

The empirical analysis presented below is based on a (disguised) data set comprising customer purchase information provided by Sun Microsystems, Inc. The data set encompasses four products and some 1,916 customers. It records the number of months (measured from the month of the earliest product launch) that elapsed before each customer bought each product. This section shows that Sun's products are not independent in several ways (for instance, all four products are servers in the same family), and we propose how to cope with this situation.

As shown in Figure 2, most of the customers did not buy all four products, and in fact about half of the customers only bought one of the products. Such sparse data would pose a challenge for many of the existing data-mining and market segmentation techniques described in §2, and in general, some form of preprocessing would be required to fill in the missing data. The separation deviation model, however, handles missing data quite routinely, without preprocessing.

As may be deduced from Figure 3, products 3 and 4 were launched together at the beginning of the observation

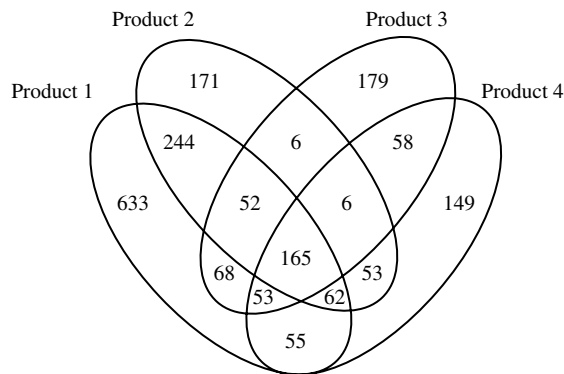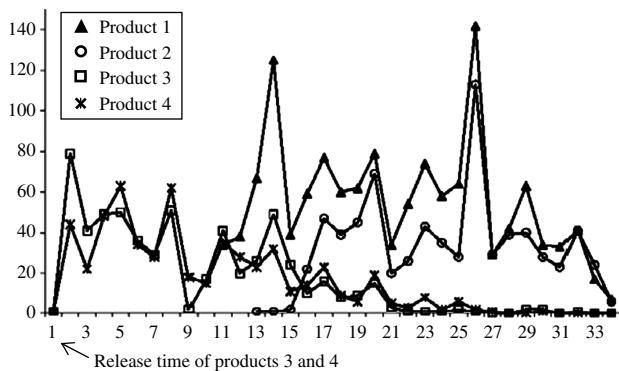**Figure 2.** Number of customers that bought each product or set of products.

Figure 3. Number of customers that bought each product per month.

Release time of products 3 and 4

period, with the launches of products 1 and 2 following, respectively, 10 and 12 months later—in fact, products 3 and 4 represent the first generation of a product line of which products 1 and 2 were the second generation, with updated and advanced features. The figure also exhibits the strong degree of correlation between product sales; peaks and valleys in the sales of all products tend to occur at the same time (this sales behavior is almost certainly due to the effect of salesforce and customer incentives that the company applied simultaneously to all products in this market). Moreover, as shown in Figure 4, most of the customers that bought products 1 and 2 did not buy either product 3 or 4.

Figure 4. Customers that bought product 1 or 2 after buying products 3 and/or 4.

(a) Product 1
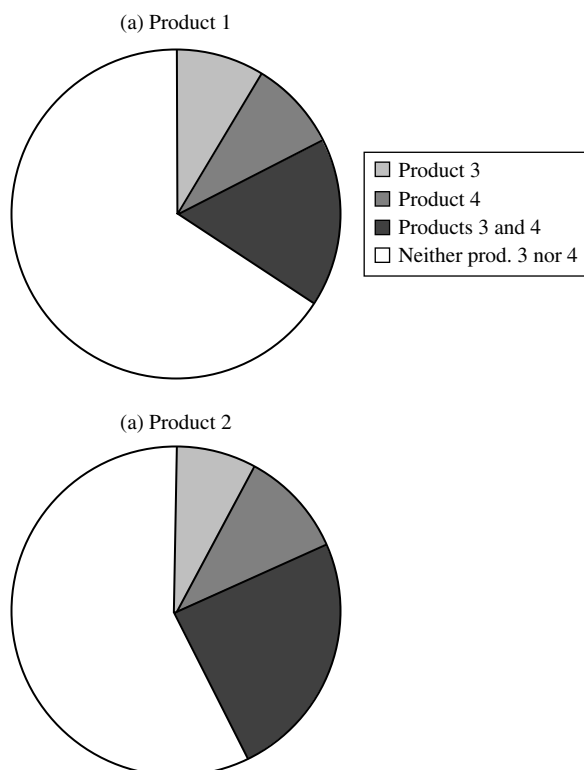
(a) Product 2

Therefore, it is reasonable to suppose that purchasers of the earlier products (products 3 and 4) exhibit greater proclivity for early adoption than purchasers of products 1 and 2 alone.

In general, the purchase times should be measured from the release date of each product. Because products 1 and 2 are the second generation of products 3 and 4, we find that the purchase times of products 3 and 4 are more significant in determining the adoption promptness. To consider this, we decided to measure all purchase times with respect to the release time of products 3 and 4. That is, we still consider the purchase times of products 1 and 2, but measure these times with respect to the launch time of products 3 and 4—as opposed to measuring these times from the launch time of the respective product.

As discussed in the previous paragraph, we believe that the purchase times of products 3 and 4 (the products with the longer life cycles) are more significant in determining the customers' adoption promptness. Consequently, we do not calibrate the data r_i^k by the length of the life cycle (as described in §3.3), and instead use as input for the SD-model the raw purchase times.

5.2. Results and Their Interpretation

In this section we demonstrate the use of the SD-model to rate Sun's customers with respect to their adoption promptness. In particular, we show that the results obtained using the SD-model agree with an intuitive interpretation of Sun's business.

Using as input Sun's data, we solved for $\mathbf{x}^{|SD|}$, which was the best performer in §4. Next, to facilitate the interpretation of the obtained results, we generated four customer classes from $\mathbf{x}^{|SD|}$. Specifically, we classified Sun's customers into the classes defined by Rogers' model of innovation diffusion (see Appendix B). That is, we segmented the customers into four classes (Vanguard, composed of innovators and early adopters; Early Majority; Early Minority; and Laggard) as follows: (1) We sorted the customers according to their rating as given by $\mathbf{x}^{|SD|}$. (2) We selected threshold values determining the boundaries between consecutive segments, so that the segments have the sizes given by Rogers's model.

Figures 5 and 6 provide an analysis of the customer segmentation in terms of customer industry and location, respectively. The bars in the figures relate the percentage of each characteristic according to customers of a particular class. Thus, in Figure 5, just under 60% of resellers are in the Vanguard, and 10% are Early Majority; whereas in Figure 6, approximately 50% of U.S. customers are in the Vanguard, and about 40% are Laggards. Broadly speaking, the results illustrated in the figures are in accord with an intuitive understanding of Sun's business. In Figure 5, for example, resellers and computer manufacturers must pass the product on to end users and thus are likely to be first in

Figure 5.   Customer classification by industry.
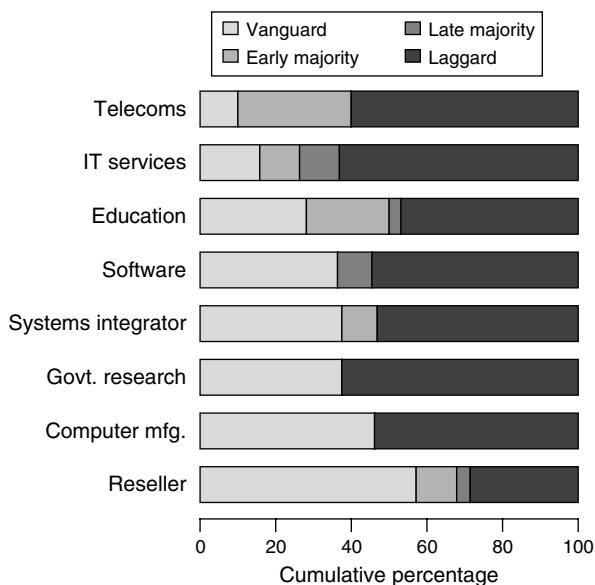
line to purchase a new model. By contrast, telecommunications utilities ("telecoms") have high fixed-capital investments and very exacting quality standards, and it is quite reasonable to see this category skewed toward the Laggard class. Figure 6 seems well grounded in the geography of Sun's markets: Because it is a U.S.-based company, one would expect a preponderance of Sun's U.S. customers to be Vanguard and Early Majority, with adoptions occurring fairly early on in the developed markets of western Europe, Australia, Japan and Canada. Less-developed markets, such as Latin America and eastern Europe, where Sun's sales and distribution infrastructure is less well established, adopt later. Overall, it appears that the classification obtained

Figure 6.   Customer classification by location.
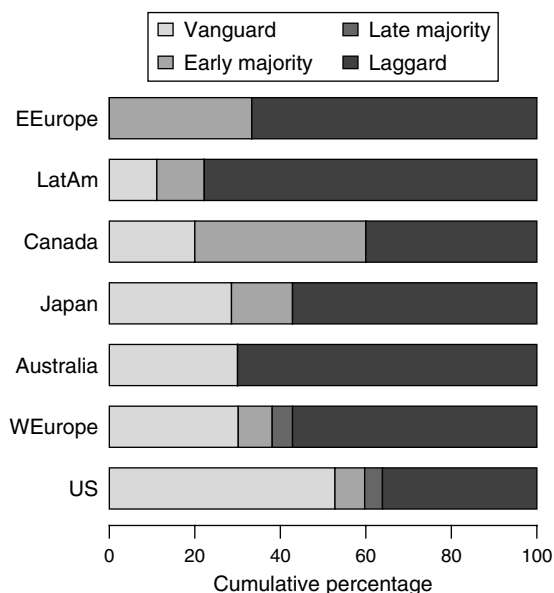
using $\mathbf{x}^{|SD|}$ does indeed characterize Sun's customers in a convincing fashion.

6. Conclusions

The proposed approach of using the (previously developed) Hochbaum's separation-deviation model (SD-model) is novel in data mining in general and customer rating in particular. It is shown here to generate valuable information on the characteristics of the customer base of an organization, and as such it is useful in managing the launch and the life cycle of a new product.

This study utilizes the SD-model to rate Sun's customers according to their adoption promptness. Using the ratings obtained, we were able to classify Sun's customers according to their adoption promptness and showed that the results provide an intuitive interpretation of Sun's business.

We further interpret the SD-model as a dimension reduction technique (specifically, to the case where multidimensional data needs to be represented/summarized in only one dimension), and we showed that the SD-model is a valuable alternative to traditional unidimensional scaling techniques. Specifically, the SD-model is shown to be better than traditional methods in terms of its scalability and its ability to deal with missing data. We thus establish that the SD-model outperforms unidimensional scaling in determining the customers' adoption promptness.

Appendix A. Kendall's Tau Rank-Correlation Coefficient

Given two rating vectors $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$, the number of *concordant pairs* is $C = \sum_{i=1}^n \sum_{j=i+1}^n C_{ij}$, where

$$C_{ij} = \begin{cases} 1 & \text{if } (a_i < a_j \text{ and } b_i < b_j) \text{ or } (a_i > a_j \text{ and } b_i > b_j) \\ & \quad \text{or } (a_i = a_i \text{ and } b_i = b_i), \\ 1/2 & \text{if } (a_i = a_j \text{ and } b_i \neq b_j) \text{ or } (a_i \neq a_j \text{ and } b_i = b_j), \\ 0 & \text{otherwise.} \end{cases}$$

Kendall's *Tau* rank-correlation coefficient is then $\tau = (4C/(n(n-1))) - 1$, and has the following properties:

1. If the two ratings imply the same ranking, then $\tau = 1$.

2. If the ranking implied by one rating is the reverse of the other, then $\tau = -1$.

3. For all other cases, τ lies between 1 and -1, and increasing values imply increasing "monotonic" agreement between the rankings implied by the ratings.

4. If the rankings implied by the ratings are completely independent, then $\tau = 0$ (on average).

Appendix B. Overview of the Customer Adoption Process

According to Rogers' now-classic model of innovation diffusion (Rogers 1962), customers may be classified, based

Figure B.1. Adoption process.
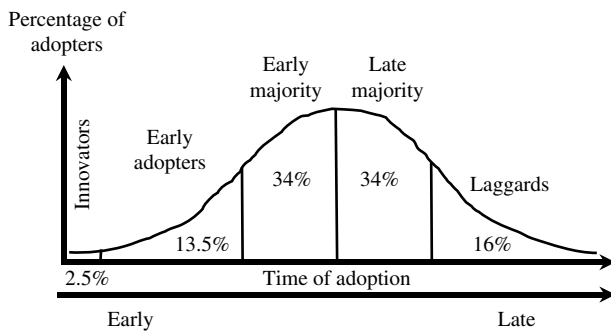
on the timing of their first purchase of a new product, as *Innovators*, who are the first to purchase the product and use it. This group of people is typically well educated, adventurous, and open to new experiences. Product purchases by those outside of this group are influenced to various degrees by the reactions of innovators. Later purchasers are essentially imitators; they buy new products because the innovators had positive reactions to them and they wish to replicate the innovators experience. *Early Adopters* begin purchasing as the innovators communicate positive responses toward a product. This group is made up of people who are inclined to try new ideas but tend to be cautious. *Early Majority* adopters are more likely to accept a new product than the average person; rarely acting as leaders, the early majority essentially imitates the behavior of the first adopters. *Late Majority* customers decide to buy only because many other customers have already tried the new product. Finally, *Laggard* Adopters are reluctant to adopt new products. Those in this group buy a new product only once this product has established a substantial track record. As seen in Figure B.1, Innovators represent only 1.5% of all customers. Thus, in §5.2 we conflate the classes Innovators and Early Adopters under the title *Vanguard*. Customers in the Vanguard class play a major role in the adoption of the innovation, because their acceptance or rejection will affect all the other groups. The method proposed in this paper aims to identify more effectively this group of customers and, therefore, increase the probability of success of a target action. Furthermore, the proposed technique identifies the other groups of customers: Early Majority, Late Majority, and Laggard. In this sense, it is a valuable instrument for the design of marketing strategies.

Acknowledgments

References

Afifi, A. A., R. M. Elashoff. 1966. Missing observations in multivariate statistics: I. Review of the literature. *J. Amer. Statist. Assoc.* **61**(315) 595–604.

Ahuja, R. K., D. S. Hochbaum, J. B. Orlin. 2003. Solving the convex cost integer dual network flow problem. *Management Sci.* **49**(7) 950–964.

Alexander, M. M., A. M. Bronstein, R. Kimmel, I. Yavneh. 2005. Multigrid multidimensional scaling. *Numer. Linear Algebra Appl.* **13**(2) 149–177.

Bartholdi, J., C. A. Tovey, M. A. Trick. 1989. Voting schemes for which it can be difficult to tell who won the election. *Soc. Choice Welfare* **6**(2) 157–165.

Brusco, M. J. 2002. Integer programming methods for seriation and unidimensional scaling of proximity matrices: A review and some extensions. *J. Classification* **19**(1) 45–67.

Brusco, M. J., S. Stahl. 2005a. Bicriterion seriation methods for skew-symmetric matrices. *British J. Math. Statist. Psych.* **58**(2) 333–343.

Brusco, M. J., S. Stahl. 2005b. *Branch-and-Bound Applications in Combinatorial Data Analysis*. Springer, New York.

Brusco, M. J., S. Stahl. 2005c. Optimal least-squares unidimensional scaling: Improved branch-and-bound procedures and comparison to dynamic programming. *Psychometrika* **70**(2) 253–270.

de Leeuw, J. 1977. Applications of convex analysis to multidimensional scaling. J. R. Barra, F. Brodeau, G. Romier, B. van Cutsem, eds. *Recent Developments in Statistics*. North Holland Publishing Company, Amsterdam, 133–146.

Fisher, R. A. 1922. The systematic location of genes by means of crossover observations. *Amer. Naturalist* **56**(646) 406–411.

Ge, Y., Y. Leung, J. Ma. 2005. Unidimensional scaling classifier and its application to remotely sensed data. *Geoscience and Remote Sensing Sympos.*, IEEE International, New York, 3841–3844.

Groenen, P. J. F., W. J. Heiser, J. J. Meulman. 1999. Global optimization in least-squares multidimensional scaling by distance smoothing. *J. Classification* **16**(2) 225–254.

Hochbaum, D. S. 2004. 50th anniversary article: Selection, provisioning, shared fixed costs, maximum closure, and implications on algorithmic methods today. *Management Sci.* **50**(6) 709–723.

Hochbaum, D. S. 2006. Ranking sports teams and the inverse equal paths problem. P. G. Spirakis, M. Mavronicolas, S. C. Kontogiannis, eds. *Internet and Network Econom., Second Internat. Workshop, WINE*, Lecture Notes in Computer Science, Vol. 4286. Springer, New York, 307–318.

Hochbaum, D. S., A. Levin. 2006. Methodologies and algorithms for group-rankings decision. *Management Sci.* **52**(9) 1394–1408.

Hochbaum, D. S., E. Moreno-Centeno. 2008. Country credit-risk rating aggregation via the separation model. *Optim. Methods Software* **23**(5) 741–762.

Hruschka, E. R., E. R. Hruschka Jr., N. F. F. Ebecken. 2003. Evaluating a nearest-neighbor method to substitute continuous missing values. *AI 2003: Adv. Artificial Intelligence* **2903** 723–734.

Huang, X., Q. Zhu. 2002. A pseudo-nearest-neighbor approach for missing data recovery on Gaussian random data sets. *Pattern Recognition Lett.* **23**(13) 1613–1622.

Hubert, L., P. Arabie, J. Meulman. 2001. *Combinatorial Data Analysis: Optimization by Dynamic Programming*. SIAM, Philadelphia.

Hubert, L., P. Arabie, J. Meulman. 2002. Linear unidimensional scaling in the L_2 norm: Basic optimization methods using Matlab. *J. Classification* **19**(2) 303–328.

Jollife, I. T. 1986. *Principal Component Analysis*. Springer-Verlag, New York.

Kosobud, R. 1963. A note on a problem caused by assignment of missing data in sample surveys. *Econometrica* **31**(3) 562–563.

Lau, K., P. L. Leung, K. Tse. 1998. A nonlinear programming approach to metric unidimensional scaling. *J. Classification* **15**(1) 3–14.

Mahajan, V., E. Muller. 1998. When is it worthwhile targeting the majority instead of the innovators in a new product launch. *J. Marketing Res.* **35**(4) 488–495.

Robinson, W. S. 1951. A method for chronologically ordering archaeological deposits. *Amer. Antiquity* **16**(4) 293–301.

Rogers, E. M. 1962. *Diffusion of Innovations*, 5th ed. Free Press, New York.

Rummel, R. J. 1970. *Applied Factor Analysis*. Northwestern University Press, Evanston, IL.

Rusmevichientong, P., S. Zhu, D. Selinger. 2004. Identifying early buyers from purchase data. W. Kim, R. Kohavi, J. Gehrke, W. DuMouchel, eds. *Proc. 10th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining*, ACM, New York, 671–677.

Talluri, K., G. van Ryzin, B. L. Bayus. 1998. An analysis of product lifetimes in a technologically dynamic industry. *Management Sci.* **44**(6) 763–775.

Torgerson, W. S. 1952. Multidimensional scaling: I. Theory and method. *Psychometrika* **17**(4) 401–419.