Buyer Targeting Optimization: A Unified Customer Segmentation Perspective

Jingyuan Yang
Department of MSIS
Rutgers University
Newark, NJ, USA
jingyyan@rutgers.edu

Chuanren Liu
Decision Sciences and MIS
Drexel University
Philadelphia, PA, USA
chuanren.liu@drexel.edu

Mingfei Teng
Department of MSIS
Rutgers University
Newark, NJ, USA
mingfei.teng@rutgers.edu

Hui Xiong
Department of MSIS
Rutgers University
Newark, NJ, USA
hxiong@rutgers.edu

March Liao
Citrix Inc.
Santa Clara, CA, USA
march.liao@citrix.com

***Abstract*—In marketing analytics, customer segmentation (clustering) divides a customer base into groups of similar individuals, while buyer targeting (classification) identifies promising customers. Both customer segmentation and buyer targeting help the business to improve marketing performances by allocating resources to the most profitable customers. Due to the heterogeneity across the customer groups, some studies have been made on combining the tasks of customer segmentation and buyer targeting for tailored marketing strategies. However, these efforts usually combine these two tasks in a simple step-by-step approach. It is still unclear how to implement these two tasks in a more integrated and optimized way, which is the research objective of this paper. Specifically, we formulate customer segmentation and buyer targeting as a unified optimization problem. Then, the customer segments are adaptively realized during the targeting optimization process. In this way, the integrated approach not only improves the buyer targeting performances but also provides a new perspective of segmentation based on the buying decision preferences of the customers. The unified customer segmentation and buyer targeting method not only quantifies the purchase tendency of a specific customer but also characterizes the buying decision behaviors at the segment level. We also develop an efficient *K-Classifiers Segmentation* algorithm to solve the unified optimization problem. Moreover, we show that the customer segmentation based on the buying decision preferences can also be consistent with the features on customer profiles. Finally, we have performed the extensive experiments on several real-world Business to Business (B2B) marketing data sets. The results show that our approach offers not only more accurate targeting of promising customers but also meaningful customer segmentation solutions with interpretable buying decision preferences for each customer segment.**

***Keywords*-Customer Segmentation, Buyer Targeting, Marketing Analytics, Optimization**

I. Introduction

Customer segmentation and *buyer targeting* are two intelligent components of the customer relationship management (CRM) systems. Specifically, customer segmentation targets on dividing the customer base into groups of individuals who share similar profiles, product needs, or marketing priorities. Customer segmentation provides better understanding of the customers' characteristics at a finer granularity level and enables differentiated marketing strategies to meet the customers' needs. On the other hand, buyer targeting identifies promising customers and allocates marketing resources on them to increase sales/profits. Finding better ways of customer segmentation and buyer targeting is essential to reduce marketing cost and boost business performances in modern marketing analytics [1, 2, 7].

Marketing professionals traditionally accomplish these two tasks in two independent steps: using a clustering method for customer segmentation and using a classification method for buyer targeting. Given the discrepancies among customer groups, the segmentation results have been leveraged to improve the classification performance for buyer targeting [2, 7, 17, 20]. For example, Chou et al. [7] proposed to first use K-Means clustering to segment customers and then build the segment-wise predictive models for better targeting the promising customers. Also, Apte et al. [2] developed individually tailored predictive models for each segment to maximize targeting accuracy in the direct-mail industry. In such a step-by-step approach, the buyer targeting (the second step) becomes dependent on the results of customer segmentation (the first step). However, the customer segmentation has to be implemented independently and can only provide limited improvements for the subsequent buyer targeting.

Is it possible to further optimize the buyer targeting performances by implementing these two tasks in a more integrated way? This is the research question we would like to answer in this paper. To this end, we investigate how to integrate these techniques in a unified optimization framework. Our key idea is to group the customers based on their decision preferences, which are quantified with the targeting models. In this way, it is possible to divide the customer base in an optimal way with respect to the targeting performance. Also, segmentation and targeting in our unified approach are intrinsically related and can be mutually supportive. To the best of our knowledge, the integration of both segmentation and targeting into a unified and optimized process is an innovative approach in solving classic marketing problems. In a more general sense, our approach unifies the clustering process and the classification modeling so that the clustering solution can best boost the classification performance. Our algorithm can also be used for other applications in addition to marketing optimizations.

More specifically, we first provide a novel mathematical formulation to integrate the customer segmentation and the buyer targeting into a unified optimization problem. Inspired by the K-Means clustering, we then develop an iterative algorithm (*K-Classifiers Segmentation*) to optimize the customer segmentation and targeting simultaneously. In comparison with K-Means clustering using the centroid as the partition and update criteria, we learn a set of targeting models and assign each customer to his/her most appropriate model. The customers assigned to the same targeting model form one customer segment, where the customers' buying decision preferences can be explained by the associated targeting model. As a result, targeting accuracy is improved due to the buying decision oriented segmentation.

To improve the interpretability and the robustness of the results, we further develop a *profile-consistent K-Classifiers Segmentation* algorithm. Indeed, using the straightforward process similar with K-Means clustering, the resultant segmentation may group customers with similar profiles into very distinct segments, which are difficult for marketing professionals to interpret. To solve the profile inconsistency issue, we exploit the Nearest Neighbor Clustering framework [5]. With the profile-consistent algorithm, the identified segmentation is consistent with not only the customer profiles but also the customer decision preferences.

Finally, we demonstrate our approach on both synthetic data set and real-world B2B marketing data sets. We use the synthetic data to better illustrate the algorithm details. The results of the real-world data sets show that the proposed approach can greatly improve the accuracy of the buyer targeting than other benchmark methods. Moreover, we validate that our approach can also offer good clustering performance comparing to K-Means clustering. In addition, the interpretation of the buying decision oriented segmentation result can help us to understand the different behaviors of the customers. The marketers can use the decision preferences of each customer segment to develop tailored marketing campaigns to attract prospects.

In summary, the contributions of our paper include:

- An innovative formal framework to integrate the customer segmentation and buyer targeting into a unified optimization problem;
- A practical profile-consistent K-Classifiers Segmentation algorithm to optimize the unified problem with interpretable solutions;
- An empirical study on real-world data, showing promising results on both better targeting accuracy and segmentation with actionable marketing implications.

Overview. The remainder of this paper is organized as follows. In Section II we provide a detailed description of our integrated framework, and in Section II-C we propose a Profile-Consistent algorithm to enhance our approach. Next, Section III reports the experimental results for both synthetic data set and real-world B2B data set. Section IV shows the related works and finally Section V concludes this work.

II. Buyer Targeting with Unified Segmentation

In this section, we develop the unified optimization framework to address the following two complementary tasks:
1) **Customer Segmentation**: The task of customer segmentation is to find a set of segments, where similar customers are grouped together. A clustering algorithm is often used for this purpose by treating each cluster as one segment.
2) **Buyer Targeting**: With a large number of customers, buyer targeting uses classification algorithms to identify the promising prospects for marketers to pursue.

In the literature, to cope with the heterogeneity of the customers, the buyer targeting models are often learned for each segment separately, and conventionally, the customer segmentation is an independent pre-step. In this work, we show that these two tasks are indeed intrinsically related and can be mutually supportive. We propose a unified framework to simultaneously segment the customers and fit the buyer targeting models. The identified segmentation is consistent with not only the customer profiles but also the customer decision preferences, which are quantified with the loss function of a targeting model.

Specifically, we unify the two tasks as an integrated optimization problem. Suppose we have a data matrix $X \in \mathbb{R}^{N \times D}$, where the n-th row x_n represents the profile features of the n-th customer. Also, we have the responses $\{y_n | n = 1, \cdots, N\}$, with $y_n = +1$ for buyers and $y_n = -1$ otherwise. We want to group the customers into K segments (clusters), $\{S_1, S_2, \cdots, S_K\}$, and learn the buyer targeting model (classifier) C_k for each segment S_k respectively. In the following, we take the linear model as the example, i.e., the decision function of C_k is of the form:

$$C_k(x) = \langle x, h_k \rangle + c_k,$$

where h_k represents the model coefficients and c_k is a constant bias.

With the linear buyer targeting models, we simultaneously group the customers and optimize the model parameters by minimizing the following *total loss*:

$$\mathcal{J}(S,C) = \frac{1}{N}\sum_{n=1}^{N} \text{loss}(x_n, y_n|C_{\ell_n}) \qquad (1)$$
$$= \frac{1}{N}\sum_{k=1}^{K}\sum_{n\in S_k} \text{loss}(x_n, y_n|C_k)$$

where ℓ_n is the segment (cluster) assignment of the n-th customer, i.e., $\ell_n = k$ if and only if $n \in S_k$. In other words, each customer-specific loss in the total loss is computed with the user's respective targeting model.

This formulation is flexible enough to incorporate different types of loss functions in different classification models. In this paper, we consider both Logistic Regression and Support Vector Machine (SVM), while other models can also be applied. To be specific, when applying Logistic Regression, we have the *logit* loss:

$$\text{loss}(x, y|C) = \log(1 + \exp(-y \cdot C(x))), \qquad (2)$$

and when applying SVM, we have the *hinge* loss:

$$\text{loss}(x, y|C) = \max\{0, 1 - y \cdot C(x)\}. \qquad (3)$$

Here, x is a customer profile, y is the corresponding response, and C is the buyer targeting model.

A. K-Classifiers Segmentation

Intuitively, the objective in Equation 1 is very similar to that of the K-Means clustering. We replace each centroid in K-Means clustering with a classifier, and we replace the distance between the centroid and a nearby point with the loss of that point in our classification model. Consequently, the problem to minimize $\mathcal{J}(S,C)$ is NP-hard. Starting with a random initialization, we use an iterative process to optimize the segmentation and targeting models (Algorithm 1). The K-Classifiers Segmentation Algorithm begins with the randomly formed (or predefined) segments. Then, the following two steps are iterated until convergence. First, in the update step, we learn a classifier C_k for each cluster S_k. Second, in the assignment step, we assign every point x_n (with response y_n) to the ℓ_n-th classifier with the minimum loss based on the specified loss function.

Although the optimization process is similar with K-Means clustering, the K-Classifiers Segmentation Algorithm does not inherit critical shortcomings of K-Means clustering. Particularly, the simple K-Means is based on Euclidean distances, and thus it is prone to generate spherical clusters with similar sizes. However, the real data sets may not satisfy these assumptions. In contrast, the K-Classifiers Segmentation Algorithm is based on a loss function, which quantifies the customer decision preferences without any assumptions on the shapes or sizes of the segments.

Algorithm 1 K-Classifiers Segmentation Algorithm.

Input: X, Y, K, $loss$.
Output: S, C.

for $n = 1, \cdots, N$ **do**
  $\ell_n \leftarrow rand\{1, \cdots, K\}$.
repeat
  #Update step:
  for $k = 1, \cdots, K$ **do**
    Learn C_k based on $\{x_n, y_n | n \in S_k\}$.
  #Assignment step:
  for $n = 1, \cdots, N$ **do**
    $\ell_n \leftarrow \arg\min_k \text{loss}(x_n, y_n|C_k)$.
until Convergence.

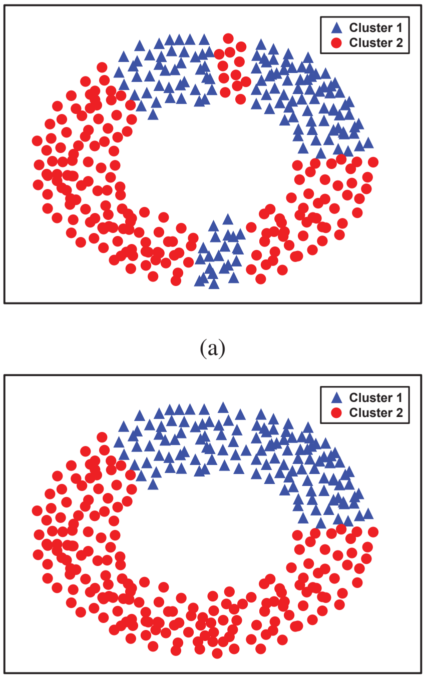
B. Profile Inconsistency Problem

The optimization in Algorithm 1 works solely with the classification loss of the data. Therefore, in some cases, it may lead to profile inconsistency. In other words, there may be the circumstance that points close to each other may end up being assigned into different clusters. For instance, Figure 1(a) shows the situation that both of two clusters (represented by blue triangle and red circle respectively) have points in the associated cluster region but belong to the other cluster (e.g. the red circles in the upper right corner and the blue triangles in the lower right corner), while Figure 1(b) shows the desired profile-consistent results. For the case of customer segmentation, the profile inconsistency means that customers with very similar feature profiles may be grouped into different segments, and consequently, it is very difficult to interpret and apply the results in practice. To cope with this challenge, we further propose profile-consistent strategy in the following section.

C. Profile-Consistent Algorithm

To improve the interpretability and the robustness of the results, now we optimize the $\mathcal{J}(S,C)$ with S_k consistent with the customer profiles $x_n \in S_k$. We adopt the *Nearest Neighbor Clustering* algorithm [5], which is very flexible and shown to produce consistent clustering solutions with arbitrary objective functions. Using the idea of Nearest Neighbor Clustering, the Profile-Consistent K-Classifiers Segmentation is provided in Algorithm 2.

In addition to settings in Algorithm 1, a new parameter M ($K \leq M \ll N$) is needed to form sub-regions in the space, and the optimization process is constrained with consistent clustering assignments for each sub-region. The sub-regions are formed with the closeness between the customers, and thus this procedure can improve the profile consistency. To be specific, we randomly select M seed points and construct the Voronoi decomposition as the sub-regions $T_1, \cdots, T_M$. Then, we start with random segment assignments for the sub-regions and iterate the update step and the assignment



(a)

(b)

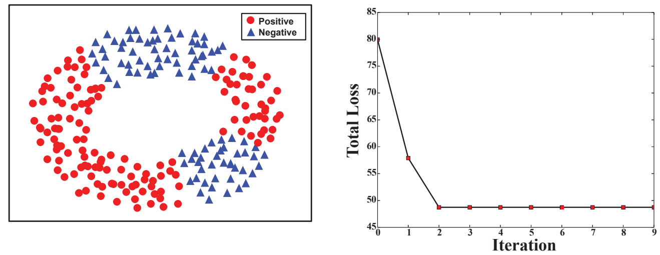
Figure 1: (a): An Example of Profile Inconsistency Case (b): An Example of Profile-Consistent Case.

step as in Algorithm 1. The only difference is that, the assignment step will group the sub-regions instead of individual points. Intuitively, we actually first identify profile-consistent sub-segments of customers, and then identify the final customer segmentation by re-allocating the sub-segments based on the targeting models. In this way, we can simultaneously learn the targeting models and identify the profile-consistent customer segments.

Algorithm 2 Profile-Consistent Algorithm.

Input: X, Y, K, M, $loss$.
Output: S, C.

Randomly select M seed points to construct the Voronoi decomposition $T_1, \cdots, T_M$.
for $m = 1, \cdots, M$ **do**
  $\ell_m^t \leftarrow rand\{1, \cdots, K\}$.
repeat
  #Update step:
  for $k = 1, \cdots, K$ **do**
    Learn C_k based on $\{x_n, y_n | n \in S_k\}$.
  #Assignment step:
  for $m = 1, \cdots, M$ **do**
    $\ell_m^t \leftarrow \arg\min_k \sum_{n \in T_m} \mathrm{loss}(x_n, y_n | C_k)$.
until Convergence.



(a) (b)

Figure 2: (a): Scatter plot of the synthetic data with two classes (red and blue). (b): The total loss at each iteration.

To further illustrate the profile-consistent algorithm, in Figure 3, we show the intermediate optimization results on a synthetic data set. Specifically, we have generated a two dimensional data set with two classes (the positive class in red and the negative class in blue), as shown in Figure 2(a). As can be seen, these two classes (red and blue) cannot be classified using just one simple linear model. However, if we partition the space into several segments, it is possible that the classes are separable within each segment. For example, the data can be divided two segments, one is the top blue points with large group of red points on left, and the other is lower blue points with small group of red points. As illustrated in Figure 2(b) and Figure 3, the profile-consistent algorithm can successfully identify the two segments and fit their respective classification models by minimizing the total classification loss, in a small number of iteration steps.

In Table I, we summarize the differences among the K-Means clustering (KM), the K-Classifiers Segmentation (KC), and the Profile Consistent K-Classifiers Segmentation (PC) with respective to three specific aspects: the initialization step, the update criteria, and the optimization objective. Again, as we mentioned in Section II-A, due to the intrinsic differences among the KM, our proposed KC and PC algorithms, the KC and PC algorithms will not inherit critical shortcomings of K-Means clustering.

D. Convergence Analysis

Similar to classical K-Means clustering algorithm, our approaches also have the property of convergence. The total cost monotonically decreases since each iteration of Algorithm 1 and Algorithm 2 necessarily lowers the cost, as shown in the following.

Let $C_1^{(t)}, \ldots, C_k^{(t)}, S_1^{(t)}, \ldots, S_k^{(t)}$ denote the classifiers and clusters at the start of the t-th iteration of both algorithms. The t-th iteration assigns each data point to the classifier with the minimum loss based on specified loss function. Therefore $\mathrm{loss}(S_1^{(t+1)}, \ldots, S_k^{(t+1)}; C_1^{(t)}, \ldots, C_k^{(t)}) \leq \mathrm{loss}(S_1^{(t)}, \ldots, S_k^{(t)}; C_1^{(t)}, \ldots, C_k^{(t)})$. Next, each cluster is reformed by the data points with corresponding classi-

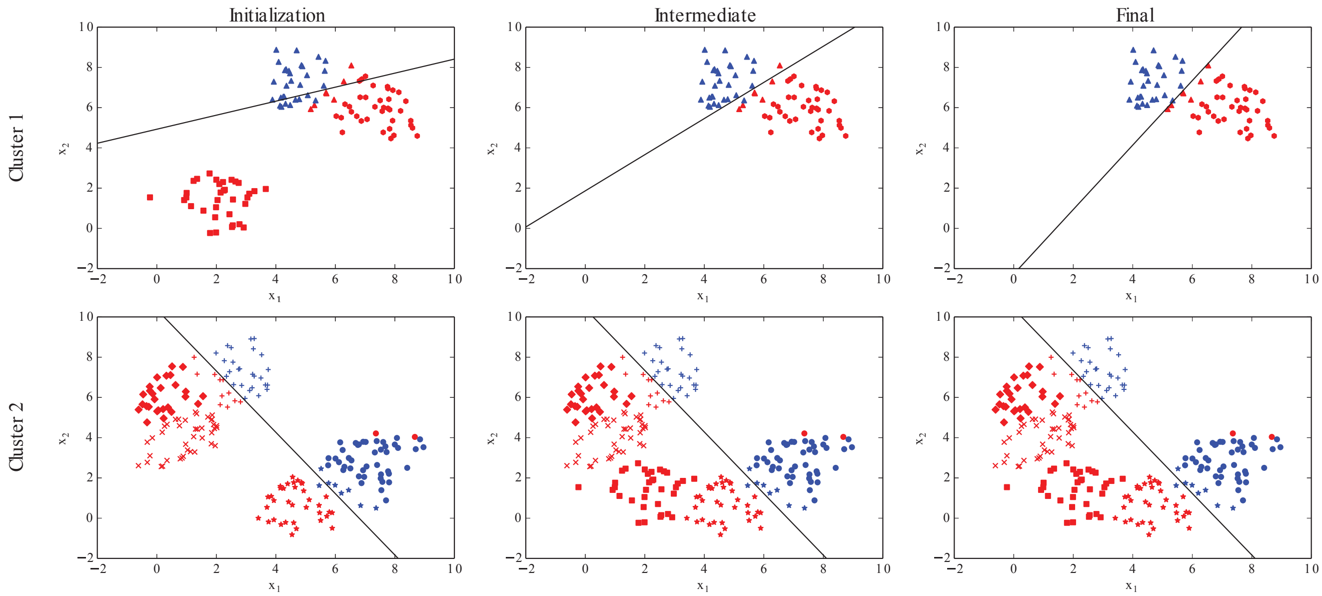


Figure 3: An illustration of Profile-Consistent K-Classifiers Segmentation Algorithm. Initialization (left): $M = 8$ sub-regions with different shapes are randomly assigned to $K = 2$ clusters (top and bottom). The classification boundary in each cluster is also plotted. Intermediate results (middle): Sub-regions are re-allocated to cluster 1 (top) or cluster 2 (bottom), according to their respective loss. Final result (right): The algorithm identified two segments both of which are linearly separable.

Table I: The Comparisons of Three Algorithms.

	K-Means	**K-Classifiers**	**Profile Consistent K-Classifiers**
Initialization	Randomly select K observations as the initial centroids.	Randomly select K observations as the initial centroids.	Randomly select M seed points to construct M sub-regions.
Update Criteria	The centroid.	The classifier learned based on all the points in the segment.	The classifier learned based on all the points in the segment.
Objective	Minimize the within-cluster distance.	Minimize the point-wise classification loss.	Minimize the classification loss for each sub-region.

fier, then $\text{loss}(S_1^{(t+1)}, \ldots, S_k^{(t+1)}; C_1^{(t+1)}, \ldots, C_k^{(t+1)}) \leq \text{loss}(S_1^{(t+1)}, \ldots, S_k^{(t+1)}; C_1^{(t)}, \ldots, C_k^{(t)})$.

III. Experimental Results

In this section, we demonstrate the effectiveness of our approach on both synthetic data and real-world B2B marketing data. All the experiments are performed on a Window 7 system with 2 CPUs (Intel i5 2.5GHz) and 8G RAM.

A. The Experimental Setup

Synthetic Data: As aforementioned, we have simulated a small two-dimensional data set with two classes represented by red and blue color, respectively, as shown in Figure 2(a). Some data characteristics of the synthetic data are summarized in Table II.

Real-world B2B Marketing Data: In this study, we obtained a B2B marketing data from a large multinational software company. Specifically, we have two sets of customers interested in different products. One is the network appliance denoted as Product A, and the other one is the desktop visualization software denoted as Product B. With "dormant" customer (no activities for six months) records removed, we have remaining $30,475$ customer records ($8,315$ for Product A and $22,160$ for Product B respectively). Each customer record includes 49 profile attributes and a binary class label to indicate buyer or otherwise.

More specifically, the B2B marketing data set includes demographic attributes, such as industry, company size, and job title information. We also have detailed customer behavior attributes (valued by interaction counts) related to customers' interactions with the company through four major types of marketing campaigns, such as *Event* related campaigns, *Offer* related campaigns, *Product Trial* related campaigns, and *Activity* related campaigns. These behavior attributes reveal meaningful insights about how the prospects behave in specific campaign activities and show their preferences of the marketing campaigns. The data characteristics are shown in Table II, and more details of the attributes are summarized in Table III.

Table II: Synthetic and Two Real-world B2B Data Sets.

Data	Size	Positive Class	Negative Class
Synthetic	300	193	107
Product A	8,315	1,680	6,635
Product B	22,160	4,232	17,828

Table III: Demographic Variables and Behavior Variables.

Demographics	Values
Company Size	Small Business, Enterprise, Unknown
Industry	Heavy Hitters, Potentials
Job Title	IT Staff, IT Manager, Executive, Researcher, Non-IT, Unknown
Behaviors	**Values**
Event	Corporate Event, Trade Show, Conference, Webinar, Seminar, Technology Preview
Offer	Official Website, Direct Mail, Email, Call Center, Search Engine, Web Advertising (third party), Social Media
Product	Product Download, Product Free Trial, Product Renewal, Product Activation, Product Training
Activity	Subscribe, Unsubscribe, Active, Inactive

Benchmark Methods: We compare our approaches with other benchmark methods using two base classifiers (LR and SVM), and we summarize all the methods as follows, where the last two methods (KC and PC) are proposed in this paper.

1) Single Classification (LR and SVM): A single general classification approach, not segment-wise classification.
2) Segment-wise Classification using K-Means clustering (SW_{LR} and SW_{SVM}): A step-by-step approach to first use K-Means clustering to do segmentation and then develop classification models for each segment.
3) K-Classifiers Segmentation Approach (KC_{LR} and KC_{SVM}): An integrated approach to use K-Classifiers Segmentation Algorithm to find out both meaningful segmentation and optimized classification models.
4) Profile Consistent Approach (PC_{LR}, PC_{SVM}): An integrated approach to use Profile Consistent Algorithm to ensure profile consistency.

B. Comparison of Targeting Performances

We apply ten-fold cross validation to evaluate the targeting/classification performance with six widely used metrics: *Accuracy*, *Precision*, *Recall*, *F-measure*, *Average Precision Score* (AP), and *Area Under Curve* (AUC) [19]. We report the average performances of the cross-validation in Table IV for both the synthetic data and the real-world B2B marketing data. From Table IV, we have the following observations:

- First, we clearly see that our PC approach outperforms other baselines under different metrics, which demonstrates the effectiveness of our framework for buyer targeting optimization. In general, PC has a much better classification performance than KC due to the profile-consistency property of PC.
- Second, PC outperforms other baselines with a more significant margin on synthetic data than on the B2B marketing data. The class ratio of the synthetic data is more balanced, and PC shows a great advantage in terms of all six measures. While for the B2B marketing data, PC achieves better results in terms of *Precision*, *Recall*, *F-measure* and AP. This may be due to that the B2B marketing data is more sparse and with relatively high imbalanced class ratio.
- Third, with utilization of either logistic regression or SVM base classifier, our approach works well and stable on all the data sets. This observation well affirms the advantage of our approach that it is very flexible and can adopt any loss functions.
- Lastly, in some cases, the segment-wise classification approach may have similar performance or even slightly outperform our approach. However, for these two classic and challenging marketing problems, our approach is an innovative attempt for enhancing the marketing performance in a new way and can get improvement from baselines. Also, it is worth noting that the greater performance improvement from KC to PC approach than from SW to PC.

C. Comparison of the Clustering Performances

In addition to targeting performance, we also compare the clustering performance of our approaches to the benchmark methods. Since there is no external clustering labels as true labels, so we measure the goodness of the clustering results using some internal clustering validation measures based on the compactness and separation.

Clustering Evaluation Metrics: $Calinski - Harabasz$ (CH) index, I index and $Silhouette$ index are used are the evaluation measures [15]. The formulas to calculate the metrics are as follows:

$$CH : \frac{\sum_i n_i d^2(c_i, c)/(NC-1)}{\sum_i \sum_{x\in S_i} d^2(x, c_i)/(n-NC)},$$

$$I : (\frac{1}{NC} \times \frac{\sum_{x\in D} d(x, c)}{\sum_i \sum_{x\in S_i} d(x, c_i) \times \max_{i,j} d(c_i, c_j)})^P,$$

$$Silhouette(S) : \frac{1}{NC} \sum_i \{\frac{1}{n_i} \sum_{x\in S_i} \frac{b(x)-a(x)}{max[b(x), a(x)]}\},$$

where D: data set; n: number of objects in D; c: center of D; P: attributes number of D; NC: number of clusters; S_i: the i-th cluster; n_i: number of objects in S_i; c_i: center of S_i; $d(x, y)$: distance between x and y; $a(x) = \frac{1}{n_i - 1} \sum_{y\in C_i, y\neq x} d(x, y)$; $b(x) = \min_{j,j\neq i}[\frac{1}{n_j} \sum_{y\in C_j} d(x, y)]$.

The CH index validates the cluster performance based on the average between- and within-cluster sum of squares. Index I (I) measures both separation and compactness in terms of the maximum distance between cluster centers and the sum of distances between objects and their cluster center

Table IV: The comparisons of targeting performances. For all methods, all parameters (if any) are empirically selected through cross-validation.

		$Accuracy$	$Precision$	$Recall$	$F-measure$	AP	AUC
Synthetic (K=2, M=8)	LR	0.6466 ± 0.07	0.7295 ± 0.05	0.7492 ± 0.14	0.7325 ± 0.07	0.8226 ± 0.03	0.5964 ± 0.08
	SW_{LR}	0.8800 ± 0.06	0.9441 ± 0.05	0.8744 ± 0.08	0.9052 ± 0.05	0.9510 ± 0.03	0.8831 ± 0.06
	KC_{LR}	0.5296 ± 0.16	0.6861 ± 0.15	0.5279 ± 0.16	0.5918 ± 0.15	0.7636 ± 0.10	0.5299 ± 0.17
	$\mathbf{PC_{LR}}$	$\mathbf{0.9499} \pm 0.05$	$\mathbf{0.9651} \pm 0.03$	$\mathbf{0.9600} \pm 0.05$	$\mathbf{0.9618} \pm 0.04$	$\mathbf{0.9759} \pm 0.02$	$\mathbf{0.9450} \pm 0.05$
	SVM	0.7039 ± 0.08	0.8272 ± 0.14	0.7692 ± 0.18	0.7686 ± 0.07	0.8751 ± 0.05	0.6746 ± 0.13
	SW_{SVM}	0.7935 ± 0.10	0.9018 ± 0.12	0.8092 ± 0.16	0.8345 ± 0.10	0.9188 ± 0.05	0.7846 ± 0.14
	KC_{SVM}	0.5873 ± 0.10	0.8630 ± 0.13	0.4834 ± 0.17	0.5897 ± 0.14	0.8444 ± 0.05	0.6367 ± 0.09
	$\mathbf{PC_{SVM}}$	$\mathbf{0.9268} \pm 0.06$	$\mathbf{0.9688} \pm 0.03$	$\mathbf{0.9200} \pm 0.07$	$\mathbf{0.9422} \pm 0.05$	$\mathbf{0.9710} \pm 0.02$	$\mathbf{0.9300} \pm 0.05$
Prod-A (K=5, M=60)		$Accuracy$	$Precision$	$Recall$	$F-measure$	AP	AUC
	LR	0.8922 ± 0.01	0.7737 ± 0.02	0.6536 ± 0.03	0.7081 ± 0.03	0.7483 ± 0.02	0.8028 ± 0.01
	SW_{LR}	0.8942 ± 0.01	0.7798 ± 0.03	0.6554 ± 0.03	0.7132 ± 0.02	0.7530 ± 0.02	$\mathbf{0.8055} \pm 0.01$
	KC_{LR}	0.5547 ± 0.13	0.2774 ± 0.20	0.5406 ± 0.06	0.3471 ± 0.14	0.4549 ± 0.12	0.5494 ± 0.10
	$\mathbf{PC_{LR}}$	$\mathbf{0.8954} \pm 0.01$	$\mathbf{0.7875} \pm 0.05$	$\mathbf{0.6578} \pm 0.04$	$\mathbf{0.7146} \pm 0.03$	$\mathbf{0.7559} \pm 0.03$	0.8054 ± 0.02
	SVM	0.8117 ± 0.05	0.5482 ± 0.13	0.5495 ± 0.09	0.5414 ± 0.09	0.5939 ± 0.08	0.7133 ± 0.05
	SW_{SVM}	0.8255 ± 0.04	0.5813 ± 0.12	$\mathbf{0.5875} \pm 0.10$	0.5738 ± 0.08	0.6257 ± 0.07	$\mathbf{0.7363} \pm 0.04$
	KC_{SVM}	0.5399 ± 0.03	0.2222 ± 0.02	0.5194 ± 0.07	0.3108 ± 0.03	0.4188 ± 0.04	0.5322 ± 0.03
	$\mathbf{PC_{SVM}}$	$\mathbf{0.8504} \pm 0.01$	$\mathbf{0.6635} \pm 0.04$	0.5165 ± 0.09	$\mathbf{0.5766} \pm 0.06$	$\mathbf{0.6383} \pm 0.04$	0.7252 ± 0.04
Prod-B (K=3, M=150)		$Accuracy$	$Precision$	$Recall$	$F-measure$	AP	AUC
	LR	0.9277 ± 0.01	0.8555 ± 0.01	0.7682 ± 0.05	0.8087 ± 0.03	0.8350 ± 0.02	0.8679 ± 0.02
	SW_{LR}	0.9281 ± 0.01	$\mathbf{0.8557} \pm 0.01$	0.7705 ± 0.05	0.8102 ± 0.03	0.8360 ± 0.02	0.8690 ± 0.02
	KC_{LR}	0.6053 ± 0.12	0.3120 ± 0.19	0.5737 ± 0.04	0.3864 ± 0.12	0.4855 ± 0.11	0.5935 ± 0.08
	$\mathbf{PC_{LR}}$	$\mathbf{0.9292} \pm 0.00$	0.8556 ± 0.01	$\mathbf{0.7771} \pm 0.02$	$\mathbf{0.8143} \pm 0.01$	$\mathbf{0.8386} \pm 0.01$	$\mathbf{0.8721} \pm 0.01$
	SVM	0.8551 ± 0.02	0.6313 ± 0.07	$\mathbf{0.7042} \pm 0.07$	0.6610 ± 0.04	0.6973 ± 0.03	0.7985 ± 0.03
	SW_{SVM}	0.8534 ± 0.03	0.6285 ± 0.07	0.6830 ± 0.09	0.6500 ± 0.06	0.6875 ± 0.05	0.7895 ± 0.04
	KC_{SVM}	0.5346 ± 0.07	0.2259 ± 0.03	0.5347 ± 0.09	0.3157 ± 0.04	0.4268 ± 0.05	0.5346 ± 0.05
	$\mathbf{PC_{SVM}}$	$\mathbf{0.8961} \pm 0.01$	$\mathbf{0.7736} \pm 0.04$	0.6871 ± 0.09	$\mathbf{0.7230} \pm 0.05$	$\mathbf{0.7616} \pm 0.03$	$\mathbf{0.8177} \pm 0.04$

respectively. The $Silhouette$ index measures the clustering validity based on the pairwise difference of between- and within-cluster distances. In addition, the larger the values of these three metrics, the better the clustering results.

Except for the single classification approaches, we compare the clustering performances of our proposed approach (KC and PC approaches) to the Segment-wise (SW) approach using K-Means clustering with the aforementioned three metrics. As the experiment results show in Table V, in general, all methods can achieve better clustering results on synthetic data set than real-world B2B data sets. This is because the real-world data sets are more complicated with higher dimensionality than the synthetic data set. Moreover, we can see that our KC_{SVM} approach performs slightly better than Segment-wise approach based on K-means clustering on the synthetic data set. While for the two real-world data sets, KC_{LR} and PC_{SVM} can achieve significant higher values of three metrics. This observation well demonstrate that the our approaches can also achieve great clustering results, and the unified clustering and classification approach can mutually benefit each other.

D. Decision Oriented Segmentation Analysis

Our approach also provides deep insights in addition to the improved targeting performances. Through the integrated process, distinct classification models for each segment are developed to capture various segment-wise buying decision preferences. Particularly, the attribute coefficients in the classification models can reflect the significance of the impact on the buying decision. Taking Product A for example, we plot the absolute values of the attribute coefficients for five segments as shown in Figure 4. The darker the cell color, the larger the value of the coefficient.

As can be seen, each segment has different and diversified sets of significant features, which reveals the different characteristics and buying preferences of the specific segment. Moreover, the set of important features of segment 1 to 5 move from left to right. In general, for customers interested in product A, most important features are among the *Offer* related campaigns and *Event* related campaigns. In contrast, the *Product* related and *Activity* related campaigns are less influential.

To further understand the decision preference of the customers, we list the top influential variables for several segments in Table VI. As can be seen, each segment shows different significant variables. 1) For example, for Segment 1 of Product A, four out of five features are the *Offer* related campaigns, namely, website advertising, the company official website and the search engine advertising. Thus, we may summarize the buying preferences of Segment 1 as *Offer Campaign Oriented* segment. 2) It is worth noting that Segment 3 is defined as *Job Title Oriented* due to the job title attributes, such as “Non IT” and “Researcher”, which indicate that in this segment customers with these job titles have a more apparent buying decision pattern.

In addition, the differences of the decision preferences exist not only among segments, but also between the two products. As we mentioned before, the buying preferences

Table V: The comparisons of clustering performances.

		$K-Means$	KC_{LR}	PC_{LR}	KC_{SVM}	PC_{SVM}
Syn	$Silhouette$	0.9178 ± 0.01	0.9270 ± 0.01	0.9226 ± 0.01	$\mathbf{0.9278 \pm 0.01}$	0.9277 ± 0.01
	CH	1021.87 ± 48.1	1099.95 ± 62.3	1037.49 ± 90.5	$\mathbf{1105.113 \pm 56.3}$	1100.80 ± 54.9
	I	4373 ± 349	4661 ± 377	4429 ± 451	$\mathbf{4839 \pm 312}$	4693 ± 350
Prod-A	$Silhouette$	0.1971 ± 0.03	$\mathbf{0.2228 \pm 0.01}$	0.2206 ± 0.01	0.2036 ± 0.00	0.2088 ± 0.01
	CH	564.13 ± 9.61	$\mathbf{572.68 \pm 8.05}$	560.83 ± 4.08	568.92 ± 6.25	570.38 ± 5.73
	I	$2.16E+106$	$\mathbf{3.76E+113}$	$2.19E+112$	$1.43E+107$	$5.51E+106$
Prod-B	$Silhouette$	0.1854 ± 0.01	0.2122 ± 0.03	0.1924 ± 0.02	0.2010 ± 0.01	$\mathbf{0.2132 \pm 0.02}$
	CH	1322.96 ± 25.8	1351.2 ± 31.5	1343.18 ± 23.62	1372.06 ± 21.75	$\mathbf{1384.78 \pm 27.48}$
	I	$3.80E+102$	$2.81E+109$	$1.03E+107$	$3.09E+108$	$\mathbf{1.51E+111}$

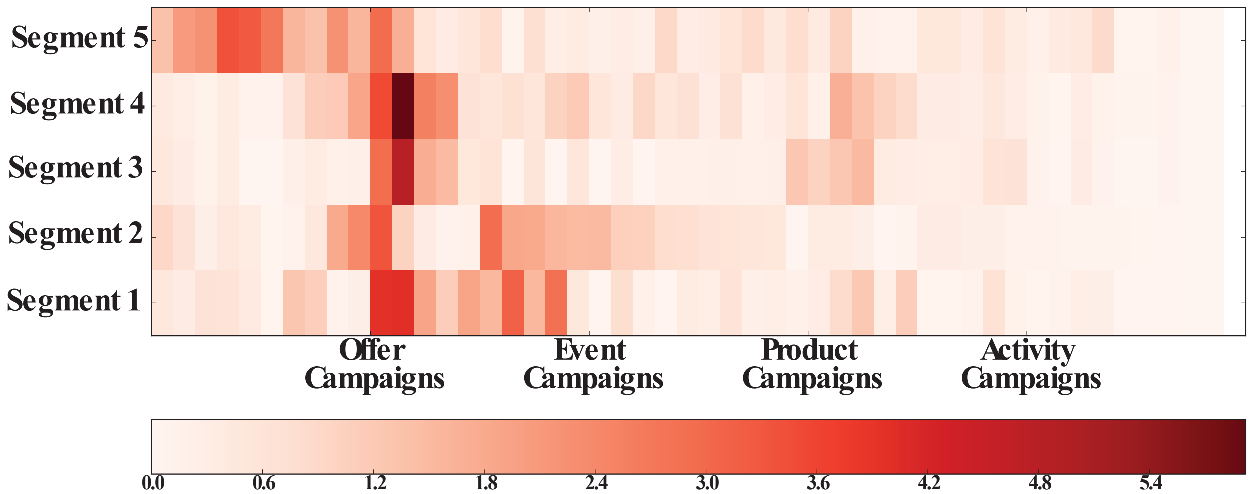
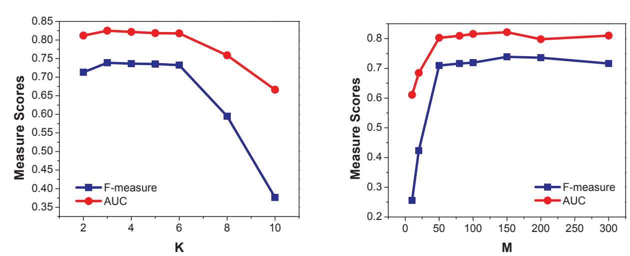


Figure 4: A heat map of the targeting coefficients of profile variables for Product A.

of Product A's segments mostly focus on the *Offer* related campaigns and *Event* related campaigns. In contrast, Product B have various kinds of buying preference patterns. For example, Segment 1 seems to be *Product Campaign Oriented* due to the majority of the features are product campaigns such as free trial, the total number of product campaign participated in, product activation, product training. Moreover, Segment 3 demonstrates the characteristic of *Activity Campaign Oriented*. These mentioned significant variables in Product B are quite different from those of segments in Product A, which indicates that the customers of these two products behave differently. The above examples show that our approach can grasp the diversified decision preferences of different segments. In summary, the results can help the marketing managers to optimize investment on more efficacious decision preference oriented campaign strategies.

E. Parameter Sensitivity

In our algorithm, there are two parameters, K and M, which represents the number of segments and the number of subregions, respectively. We fix any one of them and investigate the sensitivity of the other one in turns. For the sake of simplicity, we only show the parameter tuning experiment results of Product B data.



(a) (b)

Figure 5: (a): Impacts of The Number of Clusters on Model Performances. (b): Impacts of The Number of Sub-Regions on Model Performances.

First, we show how to decide the optimal number of clusters K. As shown in Figure 5a, we plot the classification performances with increasing number of clusters by fixing $M = 150$. As can be seen, the performances in terms of different measures ($F-measure$ and AUC) vary significantly with the different number of clusters. It is worthy noting that the performances might not increase with either smaller or larger number of clusters. The reason is that, on the one hand, a smaller number of clusters may not be enough to capture the real natural clusters, on the other hand,

Table VI: The top 5 most significant variables per segment for Product A & B.

Segment	Significant Variables	Property
	Product A	
S1	**Offer-WebAds**, **Offer-OfficialSite**, **Offer-SearchEngine**, **Offernum-total**, Evnt-Webinars	Offer Campaign Oriented
S2	**Evntnumtotal**, **Evnt-Corpevents**, **Evnt-Webinars**, **Evnt-Conferences**, Activitynumtotal	Event Campaign Oriented
S3	Evntnumtotal, **Non-IT**, Activitynumtotal, **Researcher**, Evnt-Webinars	Job Title Oriented
S4	**Offer-SocialMedia**,**Offer-CallCenter**, Executive, **Offer-Email**, **Offer-Directmail**	Offer Campaign Oriented
S5	**Evnt-Tradeshow**, **Evnt-Seminars**, **Evnt-Webinars**, Offer-Email, **Evnt-TechPrev**	Event Campaign Oriented
	Product B	
S1	**Prod-Trialfree**, **Prodnumtotal**, **Prod-Activation**, Evntnumtotal, **Prod-Training**	Product Campaign Oriented
S2	**Evnt-Corpevents**, **Evnt-Webinars**, Prod-Trialfree, **Evnt-Seminars**, **Evnt-Webinars**	Event Campaign Oriented
S3	**Activ-Subscribe**, **Activein3m**, **Activitynumtotal**, Offer-SearchEngine, **Activ-Unsubscribe**	Activity Campaign Oriented

a larger number of clusters may break the true shape of the natural clutters. Based on Figure 5a, we see that K in range of $\{3, 4, 5, 6\}$ have similar performances but $K = 3$ is a feasible choice which we achieve the optimal performance consistently in terms of these two measures. Therefore we choose $K = 3$ for this data set.

Second, the other parameter M, which controls the number of sub-regions using in the Algorithm 2 can also be chosen according to the classification performance. Intuitively, the sub-regions are formed to capture the closeness between the customers. Similarly to the parameter K, either a small or large number of M may not represent the closeness well or even make the sub-regions too trivial. To choose the optimal M, Figure 5a shows the classification performances with increasing value of M with fixed $K = 3$, where $M = 150$ gives the consistently optimal results in terms of F-measure and AUC.

IV. Related Work

Customer segmentation is one of the principal components of CRM since it helps to gain a deep understanding of customers' needs and characteristics [11, 18, 21, 23]. Many data mining techniques are gaining popularity in the market segmentation, such as CHAID decision tree [6, 10, 13], logistic regression [16], neural network analysis [4, 22], and K-Means clustering analysis [9]. In contrast, we formulate an optimized problem with advantages of providing a concrete segmentation focusing on buying decision preferences.

Another key problem in CRM is buyer targeting, that is, to identify the prospects that are most likely to become customers or most valuable to the company. Many database marketers are applying intelligent data mining tools to solve the problem, such as in [3] the authors focused on classification of online customers based on their online website behaviors, and [12] applied neural networks guided by genetic algorithms to target households. Comparing to the previous work that focus on providing a general predicting model for the total customer base, our approach provides an optimized segment-wise approach which can offer more customized and tailored strategies for each segment to improve the customer conversion rate.

Furthermore, the idea of using of segmentation to help build segment-wise prediction models has been recognized by many researchers. Several previous work [1, 2, 7, 8, 20] combined the segmentation and prediction together, and applied on the different business scenarios.

However, the problem for the existing works is that the combination of these two tasks is in a simple step-by-step way, which is difficult to theoretically guarantee the improvement for classification performance. In contrast, our work is distinguished by our development of a joint optimized classification framework, in which the two tasks are unified in a mutually supportive way.

In terms of general-purpose clustering research, this work is related to [14]. As shown, a specific cluster center should be computed for a given distance/loss metric used in the clustering process. In our case, the cluster center is modeled as a classifier to improved the overall classification performance. In other words, our work in this paper is an attempt to unify the supervised and the unsupervised learning methods.

V. Conclusion

Now we answer the question asked in the beginning of this paper: We can indeed optimally integrate the two essential marketing tasks, customer segmentation and buyer targeting, so that the customers are grouped into segments where the promising buyers can be most easily identified. In our approach, the two tasks are performed simultaneously in a unified optimization framework which combines the clustering and classification objectives. To solve the optimization problem, we developed an iterative *K-Classifiers Segmentation* algorithm, where the customer segments are formed with customers' buyer targeting models. Moreover, we showed that the segmentation results can also be consistent with the features on customer profiles. Finally, we applied our approach on both synthetic data and real-world Business-to-Business (B2B) marketing scenarios. Extensive experiments clearly validated the effectiveness of the proposed approach and its improvements in comparison with alternative methods. In addition to targeting (classification) accuracies, we showed that our approach can provide interpretable customer segmentation solutions and reveals new marketing insights.

References

[1] Evgeny Antipov and Elena Pokryshevskaya. Applying chaid for logistic regression diagnostics and classification accuracy improvement. *Journal of Targeting, Measurement and Analysis for Marketing*, 18(2):109–117, 2010.

[2] Chidanand Apte, E Bibelnieks, Ramesh Natarajan, E Pednault, Fateh Tipu, Deb Campbell, and Bryan Nelson. Segmentation-based modeling for advanced targeted marketing. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 408–413. ACM, 2001.

[3] Dimitris Bertsimas, Adam J Mersereau, and Nitin R Patel. Dynamic classification of online customers. In *SDM*, pages 107–118. SIAM, 2003.

[4] Derrick S Boone and Michelle Roehm. Retail segmentation using artificial neural networks. *International journal of research in marketing*, 19(3):287–301, 2002.

[5] Sébastien Bubeck and Ulrike von Luxburg. Nearest neighbor clustering: A baseline method for consistent clustering with arbitrary objective functions. *The Journal of Machine Learning Research*, 10:657–698, 2009.

[6] Jan Roelf Bult and Tom Wansbeek. Optimal selection for direct mail. *Marketing Science*, 14(4):378–394, 1995.

[7] Paul B Chou, Edna Grossman, Dimitrios Gunopulos, and Pasumarti Kamesam. Identifying prospective customers. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 447–456. ACM, 2000.

[8] Yong Ge, Hui Xiong, Wenjun Zhou, Siming Li, and Ramendra Sahoo. Multifocal learning for customer problem analysis. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):24, 2011.

[9] Mihajlo Grbovic, Nemanja Djuric, and Slobodan Vucetic. Supervised clustering of label ranking data. In *SDM*, pages 94–105. SIAM, 2012.

[10] Dominique Haughton and Samer Oulabi. Direct marketing modeling with cart and chaid. *Journal of Interactive Marketing*, 11(4):42–52, 1997.

[11] Jedid-Jah Jonker, Nanda Piersma, and Rob Potharst. A decision support system for direct mailing decisions. *Decision support systems*, 42(2):915–925, 2006.

[12] YongSeog Kim, W Nick Street, Gary J Russell, and Filippo Menczer. Customer targeting: A neural network approach guided by genetic algorithms. *Management Science*, 51(2):264–276, 2005.

[13] Nissan Levin and Jacob Zahavi. Predictive modeling using segmentation. *Journal of Interactive Marketing*, 15(2):2–22, 2001.

[14] Chuanren Liu, Tianming Hu, Yong Ge, and Hui Xiong. Which distance metric is right: An evolutionary k-means view. In *SDM*, pages 907–918. SIAM, 2012.

[15] Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, Junjie Wu, and Sen Wu. Understanding and enhancement of internal clustering validation measures. *IEEE transactions on cybernetics*, 43(3):982–994, 2013.

[16] John A McCarty and Manoj Hastak. Segmentation approaches in data-mining: A comparison of rfm, chaid, and logistic regression. *Journal of business research*, 60(6):656–662, 2007.

[17] Vicki G Morwitz and David Schmittlein. Using segmentation to improve sales forecasts based on purchase intent: Which" intenders" actually buy? *Journal of Marketing Research*, pages 391–405, 1992.

[18] Eric WT Ngai, Li Xiu, and Dorothy CK Chau. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2):2592–2602, 2009.

[19] Tan Pang-Ning, Michael Steinbach, Vipin Kumar, et al. Introduction to data mining. In *Library of Congress*, 2006.

[20] Thomas Reutterer, Andreas Mild, Martin Natter, and Alfred Taudes. A dynamic segmentation approach for targeting and customizing direct marketing campaigns. *Journal of Interactive Marketing*, 20(3):43–57, 2006.

[21] Stephen L Vargo and Robert F Lusch. Evolving to a new dominant logic for marketing. *Journal of marketing*, pages 1–17, 2004.

[22] A Vellido, PJG Lisboa, and K Meehan. Segmentation of the on-line shopping market using neural networks. *Expert Systems with Applications*, 17(4): 303–314, 1999.

[23] Peter C Verhoef, Penny N Spring, Janny C Hoekstra, and Peter SH Leeflang. The commercial use of segmentation and predictive modeling techniques for database marketing in the netherlands. *Decision Support Systems*, 34(4):471–481, 2003.