PAPER • OPEN ACCESS

Multiple resource allocation for precision marketing

To cite this article: Siyu Zhang *et al* 2020 *J. Phys.: Conf. Ser.* **1592** 012034

Multiple resource allocation for precision marketing

Siyu Zhang[1, 2, 3], Peng Liao[2], Heng-Qing Ye[2] and Zhili Zhou[1]

[1]School of Management, Xi'an Jiaotong University, Xi'an, China
[2]Faculty of Business, Hong Kong Polytechnic University, Kowloon, Hong Kong

[3]Email: henecia33@stu.xjtu.edu.cn

Abstract. In the precision marketing of a new product, it is a challenge to allocate limited resources to the target customer groups with different characteristics. We presented a framework using distance-based algorithm, K-Nearest-Neighbour, and support vector machine to capture customers' preference towards promotion channel. Additionally, on-line learning programming was combined with machine learning strategies to fit a dynamic environment, evaluating its performance through a parsimonious model of minimum regret. A resource optimization model was proposed using classification results as input. In particular, we collected data from a loan agency that offers loans to small business merchants. Our sample contained 525,919 customers who will be introduced to a new financial product. By simulating different scenarios between resources and demand, we showed an up to 22.42% increase in the number of expected merchants when K-NN was performed with optimal resource allocation strategy. Our results also show that K-NN is the most stable method to perform classification, and that distance-based algorithm has the most efficient adoption with on-line learning.

1. Introduction

1.1. Problem description

Precision marketing is an essential part of new product promotion as high-efficiency marketing captures a large number of potential customers quickly with a rational cost of promotion resources. With the accelerating pace of economic globalization and fiercer market competition, many companies are under pressure on making the right decision in best promoting their products and sell them to the right customers. An effective marketing strategy helps suppliers quickly reach their target customers, minimize goods in stock, satisfy customer needs, and ultimately maximise profit.

One motivation for conducting this research is to understand the relationship between customer features and product features so that we can map customers to the right products. According to Rust et al. [1], customers who have bought a similar product previously are likely to buy the newly launched product. For example, Yamaha extended musical instruments from organs to pianos to guitars and promoted to similar customers. With limited promotion resources, the allocation of different types of resources to receive the maximum number of product buyers/users becomes an issue. For instance, face-to-face marketing is the most effective way for product promotion while it consumes most resources, followed by promotion by phone, promotion through e-mail or other message channels. We classify all potential customers into several types to allocate different resources. In this case, a promotion activity with limited resources under uncertain demand in achieving high promotion effectiveness becomes an optimised problem, i.e., with the largest number of customers buying the new product ultimately.

We present a framework to deal with decision making in precision marketing, including customer classification and optimization of resource allocation, together with on-line learning through marketing feedback. To illustrate our framework in detail, we presented a case study of a newly launched financial product focusing on small business merchants. Our target customers are merchants who are in urgent need of funds with a highly sensitive interest rate, which differs from other financial products in its special settlement period and attractive rate. The way chosen to do promotion by the loan agency is critical since merchants have different preferences in promotion channels. It has been recognized that precision marketing has become a key means of generating users and has become increasingly important in attracting customers quickly as customers become better informed about all kinds of products in the market.

1.2. Key results

As to customer classification, we adopted three classic methodology related to machine learning, distance-based method, K-Nearest-Neighbour (K-NN), and support vector machine (SVM) over multiple heterogeneous features among potential customers. From the perspective of running time, distance-based method used the shortest time to complete the classification process; K-NN was most stable in terms of predicted accuracy although it cost twice as much time as that of the distance-based method. SVM showed a slightly higher predicted accuracy than that of distance-based method, but took a much long time to do classification due to finding its convex optimization results. Furthermore, as observations made in the training sample were all users who accepted a product that was very similar to the newly launched product, we had no information about customers who were uninterested in this new product. To deal with this issue, on-line learning strategy was proposed. We found that distance-based algorithm always reached its stable state after one episode, while K-NN and SVM had a slower speed when learning. Classification results were then taken as input parameters of resource optimization for allocation of resources. Experiments were also conducted to compare the optimal resource allocation with the marketing strategies currently adopted by the loan agency. Our simulation results showed that the higher predicted accuracy one algorithm yields, the greater the increase in final expected users, thus, the better the allocation proposal. Among the three classification algorithms, K-NN outperformed others with an 22.42% increase in final expected customers. With more expected customers wanting to try new products under the scenario of limited promotion resources, the corresponding classification methodology together with optimal resource allocation plan make an improvement towards the promotion strategy that the company adopts at present.

The rest of this paper is organized as follows. Section 2 introduces related works on both classification algorithms and resource allocation. Section 3 describes the proposed framework and data mining methodologies in detail. Section 4 presents an empirical study to further implement theoretical model, with analysed results. Finally, section 5 concludes our contributions and offers some suggestions for companies who would like to adopt precision marketing when launching a new product.

2. Literature review

One common way to do classification is through machine learning, which has been employed in many fields and is considered to be a powerful tool for data-mining. Business researchers have adopted methods such as decision tree, support vector machine to solve classification problems [2, 3]. Galindo and Tamayo developed a specific model to estimate credit risk assessment with a focus on credit risk for mortgage-loans based on the study of error curves, using predicted error rate for the comparison of different algorithms, including Probit, Classification and regression tree (CART), Neural Networks and K-Nearest-Neighbor [4]. Zhang and Hardle introduced the Bayesian Additive Classification Tree (BACT), a nonlinear classification method extending the Bayesian Additive Regression Tree (BART) for credit risk modeling and found that BACT is a serious competitor to logit model, CART, support vector machine (SVM), random forest and gradient boosting [5]. The models mentioned above mainly conduct research to do classification as a financial technology to recognize the credit level. However, they do not provide any evidence about the degree of improvement they could make to benefit financial business.

Classification problem using machining learning in other domains has also raised much attention. Turney handled a review classification problem (recommended: thumbs up; not recommended: thumbs down) with a simple unsupervised learning algorithm by identifying adjectives or adverbs in the review [6]. Specially, many research focus on algorithms for learning features from labeled data. Cui, Wong and Lui established a Bayesian network to estimate consumer behavior to improve the performance of marketing operations, and found that Bayesian networks have distinct advantages over other methods such as neural networks, decision tree and latent class regression in predicted accuracy [7]. One related research by Lin who put forward an innovative approach of semi-supervised geographic information in clustering retail customers aimed to discover useful customer patterns for marketing strategy, and the results showed reasonable clusters on real customer profile [8]. Kim developed a model of consumer response by combining genetic algorithms to select predictive demographic variables with artificial neural networks in order to identify prospective households and for customer targeting, and showed strong dependencies between model specification and managerial decision making [9]. Marcus introduced the concept of Customer Value Matrix for customer segmentation which suited small retail and service businesses [10]. Moreover, according to Hochbaum, managerial decision-making was made with the help of separation-deviation model to rate customers according to their proclivity for adopting products [11]. Several supervised learning or unsupervised learning methods have been proposed in the general literature for classification, but little empirical comparison of machine learning algorithm with traditional statistical algorithms has been made [12].

Our research was conducted via an empirical study and an optimization model with reference to Kim, who has conducted case studies over 190,000 hospitalizations across 15 hospitals to quantify the cost of denied Intensive Care Units (ICUs) admission, aiming at evaluating the performance of various admission strategies to hospital' ICU [13]. While Kim focused on a queueing system in hospital modelling by a traditional statistical method of probit, our study dealt with a resource allocation problem under the guidance of machine learning algorithms to recognize potential merchants. Wang organized his research with an empirical study, and developed a hospital quality model to evaluate hospital quality by constructing distance based instruments to correct potential selection bias in care allocation after summarizing the empirical setting and data [14]. Many scholars have adopted various methods ranging from traditional linear methods to machine learning to solve a variety of problems. We consider a dynamic scenario with uncertain demand and limited promotion resources to find an optimized solution to get maximum expected users, whose preference in promotion channel were simulated by several machine learning algorithms.

Literature related to finding the trade-off between supply and demand was further examined. Rudin and Vahn studied a newsvendor decision-maker to make a sensible ordering decision according to past information about various features of the demand. They proposed two tractable algorithms, one with a small feature-observation ratio and the other a large one. Further investigation showed that the custom-designed, feature-based algorithms yielded substantially lower cost than several main benchmarks known in the literature [15]. However, research on resource allocation in the literature mostly estimate demand of customers from past ordering information in general, while our research proposed a preference towards promotion channel dealing with a totally unknown demand of potential customers. Furthermore, our research integrates the forecasting of preferences towards promotion channel and resource allocation, rather than solving them separately as typically done.

Previous studies focused on marketing resources related to budget decisions indicated that profit improvement from better allocation across products or regions is much higher than that from improving the overall budget [16]. In fact, heuristic methods are usually used by marketing practitioners in determining marketing budget although a great deal of research have studied budget questions yet [17]. They focused on optimizing the budget for a product in a static environment, while we tried to solve the allocation problem under the consideration of uncertain market demand, i.e., the dynamic preference of customers.

In reality, some customers will not accept a newly launched product despite its promotion. To further analyse customers with no interest in the new product, we conducted an on-line convex programming algorithm to fit this situation. According to Zinkevich, on-line learning policy makes a decision before it observes the cost function [18]. Since a training set focuses on customers who have bought a similar product before instead of those with little or no interest in the new product, machine learning does not take highly loyal customers of other brands into consideration and thus, algorithms classify all potential customers with the same label of training data, resulting in all customers getting promoted, somehow a waste of resources. On-line learning separated all test observations into several rounds, i.e., getting training set updated with test feedback from the last round, and thus, adjusting classified strategies. We generate customers with a new label, i.e., of no interest, after the first round. Instead of promoting only one round, on-line learning fits an uncertain environment well with its updating adjustment.

For model implementation, data were collected from an acquiring agency who provides loans to its acquired merchants. Three classic methods related to machine learning were presented: distance-based algorithm, K-NN, and SVM over multiple different heterogeneous features for classification. We first tested whether these algorithms were stable enough when data followed distributions including Normal distribution, Uniform, Exponential, Poisson, T-test, Weibull, Logit and Lognormal. We further applied on-line learning strategy with machine learning algorithms to exclude customers who are not interested in our new product at all. Classification performance was examined by the minimum regret model, and optimization model took classification results as the input parameters later on. For comparison, we also conducted an experiment with resource allocation in the same way as the acquiring enterprise uses currently. With better optimization results, i.e., with more customers trying our new financial products under limited promotion resources, we conclude the corresponding classification method together with resource allocation plan to improve the current promotion strategy adopted by the acquiring agency to some extent.

3. Framework for precision marketing

This section describes the methodology employed in our research. The main process of precision marketing relies on the relationship between customers and products, upon which we establish data warehouse. Data layer is consisted of three databases (customer database, product database and relationship database), and is responsible for receiving real-time data, preprocessing it as well as loading data warehouse for the next layer of data analysis. The analysis layer includes customer classification and resource allocation, which are the core of the entire marketing strategy. Our classification methodology includes distance-based algorithm, K-NN and SVM to identify the preference of customers towards promotion channel. Resource allocation uses optimization models to formulate the relationship between customers demand and limited resources. We tested customers belonging to 8 distributions separately without loss of generality. The analysis layer provides marketing solutions for decision-making layer, and also obtains feedback from the decision-making layer for adjustment. The decision-making layer uses strategies generated from the analysis layer to return feedback to both data layer and analysis layer. Figure 1 illustrates our framework to implement precision marketing.

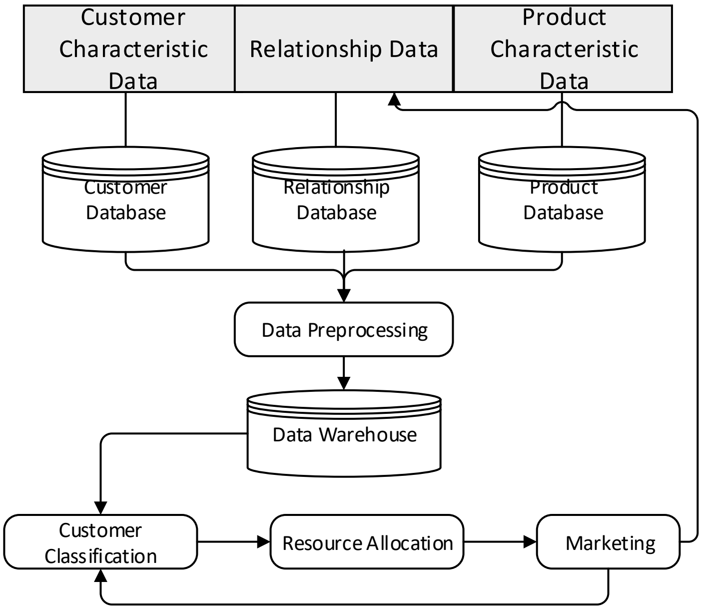


Figure 1. Framework for precision marketing.

3.1. Customer classification

In this part, we present the classification algorithms used in our research. Defining customers in the training set as $S_1, s_i \epsilon S_1, i = 1, \ldots m$, and potential customers as $S_2, s_j \epsilon S_2, j = 1, \ldots, n$, we standardize two samples and modelling according to the different scenarios below.

3.1.1. Training sample with No-label of promotion channel. When the promotion channel is unknown, denote central point s_c among training customers as $s_c = \frac{1}{m}\sum_{i=1}^{m} s_i$. We formulate distance d_j between a potential customer s_j and s_c as $d_j = \sqrt{(s_j - s_c)^2}$. Customers with smaller distance are more likely to belong to the same class. In terms of resource allocation, we consider closer customers should be first promoted. With only one type of resources, allocate all resources to customers with the minimum distance; when several types of resources exist, allocate different types of resources to customers according to the distance with the stronger promotion efficiency given to the closer customer, for example, face-to-face promotion, followed by promotion by phone, then e-mail or message.

3.1.2. Training sample with promotion channel and several. Types of Promotion Resources. In this part, we present three methods for supervised learning: distance-based method, K-NN and SVM. Resource allocation strategies will be introduced later on.

1) Distance-based Method

Based on the Euclidean distance, customer classification through distance-based algorithm could quickly identify the class a customer belongs to. Labels in training set indicate the promotion preference of customers. We split training set S_1 into p subset with S_k representing the kth class. Given s_{ck} of the kth class as an example, denote central points of each class, as

$$s_{ck} = \frac{1}{m}\sum_{i \in S_k} s_i \, , S_k \subset S_1, \tag{1}$$

where m represents the number of customers in class k. We assign a class label with minimum distance to customers, taking class $y_j(k)$ of customer j as an example,

$$y_j(k) = \arg\min_{k=1,\ldots,p} \sqrt{(s_j - s_{ck})^2}. \tag{2}$$

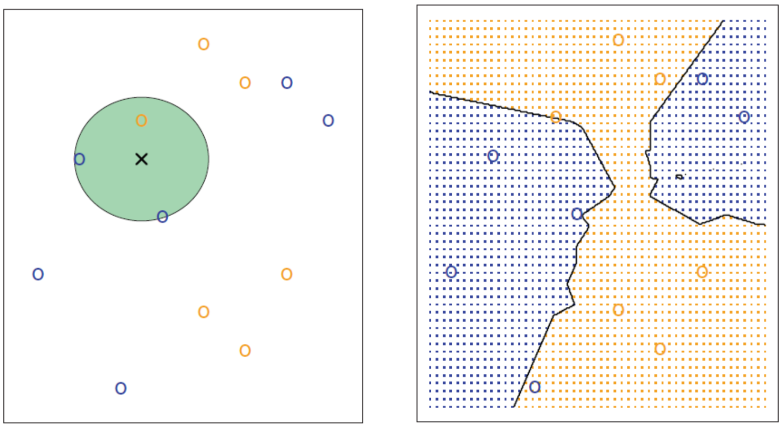
2) K-Nearest-Neighbour

The classification through K-NN resets sample variables into (X, Y), with X describing customer characteristics such as age and Y stands for the label of promotion channel like face-to-face. As the exact conditional distribution of Y given X is unknown, perform classification according to K customers in the training sample who are closest to the potential customer, and consider the highest probability of which class the potential customer belongs to on the basis of which class the K customers are in.

We provide mathematical description with reference to James [19]. Given positive integer K and one potential customer x_0, K-NN classifier identifies the K customers that are closest to x_0, represented by N_0. K-NN then estimates the conditional probability for class j as the fraction of points in N_0 whose class label equals j:

$$Pr(Y = j|X = x_0) = \frac{1}{K}\sum_{i\in N_0} I\,(y_i = j). \tag{3}$$

K-NN applies Bayes rule and classifies potential customer x_0 to the class with the largest probability. As to the selection of K, cross-validation is considered the most common method. To further explain K-NN, we present a simple situation with six blue observations and six orange observations as shown in Figure 2, with $K = 3$. Figure 2a shows a test observation at which a predicted class label is depicted as a black cross. The three closest points to the test observation are identified, and it is predicted that the test observation belongs to the most commonly occurring class, i.e., blue in this case. Figure 2b shows the K-NN decision boundary for this example in black. The blue grid indicates the region in which a test observation will be assigned to the blue class, and the orange grid indicates the region in which it will be assigned to the orange class.



(a) K-NN layout (b) Decision boundary

Figure 2. The K-NN approach, using K = 3.

3) Support Vector Machine

Support vector machine (SVM) is a special case of support vector classifier, an extension resulting from enlarging feature space in a specific way, using kernels. Take n training customers for example:

$$(\vec{x}_1, y_1), \dots (\vec{x}_n, y_n),$$

where $\vec{x}_i$, vectors in p dimensional space, describes customer characteristics such as gender while $y_i = -1$ or $y_i = 1$, representing the promotion preference of $\vec{x}_i$. SVM aims at splitting all the observations of $y_i = 1$ and $y_i = -1$ by a "maximum margin hyperplane", which is defined so that the distance between hyperplane and the nearest point $\vec{x}_i$ from either group is maximized.

For a certain hyperplane, $\vec{x}$ in that space must satisfy the following constraint:

$$\vec{\omega} \cdot \vec{x} - b = 0, \tag{4}$$

where $\vec{\omega}$ is the normal vector (not necessarily normalized) to the hyperplane and parameter $\frac{b}{\|\vec{\omega}\|}$ determines the offset from the hyperplane towards $\vec{\omega}$. We classify a potential customer based on which side of the maximal margin hyperplane it lies. To extend SVM to cases in which data are not linearly separable, we introduce the hinge loss function:

$$max(0,1-y_i(\vec{\omega}\cdot\vec{x}_i-b)). \tag{5}$$

When the constraint above is satisfied, the function is zero, which indicates $\vec{x}_i$ lies on the correct side of the margin. For data on the wrong side, the value of the function is proportional to the distance from the margin. We then hope to minimize

$$[\frac{1}{n}\sum_{i=1}^{n}\max\big(0,1-y_i(\vec{\omega}\cdot\vec{x}_i-\mathrm{b})\big)+\lambda(\|\vec{\omega}\|)^2]. \tag{6}$$

where the parameter λ determines the trade-off between increasing the margin size and ensuring that $\vec{x}_i$ lies on the correct side.

So far, our discussion has been limited to the case with two-class setting, yet marketing promotion may sometimes take several actions. In the more general case of multiple classes, the one-versus-one and one-versus-all approaches are considered to be an extension of SVM. One-versus-one constructs $\binom{K}{2}$ SVMs, each of which compares a pair of classes. For example, one such SVM might compare the kth class, coded as +1, to the $k'th$ class, coded as -1. We classify one potential customer using each of the $\binom{K}{2}$ classifiers, and tally the number of times that this potential customer is assigned to each of the K classes. Final classification is performed by assigning customer to the class to which it was most frequently assigned in the $\binom{K}{2}$ pairwise classifications. The one-versus-all approach fits K SVMs when applying SVM in the case of $K>2$ classes, i.e., each time comparing one in the K classes to the remaining $K-1$ classes. Customers are assigned to the class with the highest function value as this amounts to a high level of confidence that a customer belongs to the current class rather than to any other classes.

3.1.3. On-line learning. On-line learning focuses on problems belonging to a sequence of convex programming, each with the same feasible set but different cost function. Decisions have to be made before cost function is observed, usually when dealing with minimizing error on-line. For example, one makes a prediction of an unlabeled preference of a customer, and then a label is assigned to the customer. After that, we receive some error based on how divergent the label given is from the true label. Zinkevich applied gradient descent called greedy projection for general convex functions as described in the following way [18]. Select an arbitrary customer $x_1\in S$ and a sequence of learning rates $\eta_1,\eta_2,\ldots\in\mathbb{R}^+$. In round step t, after receiving a cost function, select the next choice x^{t+1} according to

$$x^{t+1}=P(x^t-\eta_t\nabla c^t(x^t)),$$

where $P(y)=arg\,min_{x\in S}d(x,y)$ is considered as the projection and $d(x,y)$ the distance between customer x and customer y. Also, it is assumed that the feasible set S is nonempty, bounded and closed and that the cost function c^t is differentiable for all t.

We combine on-line strategy with machine learning to formulate situation when we misunderstand customers with little or no interest in newly launched products since the training set in hand usually do not have this kind of label. Feedback are collected for better adjustment in the next round from the market.

3.1.4. Evaluating alternative classification algorithms. One primary objective of our research is to maximize expected buyers/users after promotion, an essential step in comparing different classification strategies. Customers would only become product buyers/users when the right resource is allocated to the right customers. We adopt parsimonious model of minimum regret to examine how estimated measures are utilized. We leverage our estimation results to calibrate a simulation model to allow us to compare the outcomes across different machine learning methodologies. The minimum regret algorithm defines the regret by calculating the differences of expected users between the best fitted allocating proposal and the other machine learning methodology. Later we compare the performance of multi methods with this baseline. In customer classification, simulation results performed among several probability distributions provide the expected regret as well as predicted accuracy. Based on classification results, the resource allocation model in the next section will seek for a minimum expected regret to show the extent in which different classification solutions together with our optimal allocation strategies improve the number of users ultimately. We translate the minimum regret method into the following with studies conducted by others [20, 21, 22].

Let $\chi \subset R^D$ be the bounded domain of all potential customers and $f: \chi \rightarrow \mathbb{R}$ be the corresponding classification function with its value standing for the class label. Our aim is to find the maximum x^* of f on χ. Assume a probability measure $p(f)$ over the space of function $f: \chi \rightarrow \mathbb{R}$. Based on this $p(f)$, we are ultimately likely to select an $\tilde{x}$ with minimum regret $R_f(\tilde{x})$. Define the expected regret ER of selecting parameter x under $p(f)$ as:

$$ER(p)(x) = \mathbb{E}_{p(f)}[R_f(x)] = \mathbb{E}_{p(f)}[\max_x f(x) - f(x)]. \quad (7)$$

3.2. *Resource allocation*

Assume the total amount of resources for promotion is R. To simplify multiple allocation process, our study only consider two types of promotion channels, and face-to-face promotion is referred to high efficiency while promotion by phone is referred to low efficiency. Different marketing channels require different resources and yield different efficiency. We assume face-to-face promotion consumes m unit of resources while promotion by phone consumes n unit of resources ($m > n$). For further analysis, we assume $m/n = \beta$. Allocate $\alpha * R$ to promote face-to-face and the rest $(1-\alpha) * R$ to promote by phone. As α ranges from 0 to 1, the strategy moves from putting all resources on promotion by phone to putting all to promotion face-to-face. More realistically, consider customers in the face-to-face promotion demand have a total number of ω_0 and the number of customers preferred to be promoted by phone is ω_1. To maximize the number of total buyers/users, we state the expected users (EU) as follows:

$$EU = P_0 * min\{\alpha * \frac{R}{m}, \omega_0\} + P_1 * min\{(1-\alpha) * \frac{R}{n}, \omega_1\}, \quad (8)$$

where $\alpha \in (0,1)$, P_0 represents the accuracy of face-to-face promotion while P_1 stands for the accuracy of promotion by phone. Take accuracy table as an standard example to calculate P_0 and P_1. When customer classification is only applied with machine learning, the item "of no interest" will be excluded in Table 1, then $P_0 = (x_{11} + x_{12})/(x_{11} + x_{12}) = 1$, $P_1 = x_{22}/(x_{21} + x_{22})$; when the performance of customer classification is consistent with both machine learning and on-line learning, then $P_0 = (x_{11} + x_{12})/(x_{11} + x_{12} + x_{13})$, $P_1 = x_{22}/(x_{21} + x_{22} + x_{23})$. For optimization, rewrite our allocation problem as:

$$\max\{P_0 * min\{\alpha * \frac{R}{m}, \omega_0\} + P_1 * min\{(1-\alpha) * \frac{R}{n}, \omega_1\}\}. \quad (9)$$

Table 1. An example of classification results.

actual / predicted	face-to-face	phone	of no interest
face-to-face	x_{11}	x_{12}	x_{13}
phone	x_{21}	x_{22}	x_{23}
of no interest	x_{31}	x_{32}	x_{33}

In reality, an imbalance between customer demand and limited promotion resources always exists. Thus, consider four scenarios:

1) Some customers still do not receive any promotion even if all resources are consumed. We select potential customers randomly under resource constraints. Formulate the optimal problem as

$$\begin{aligned} \max \quad & P_0 * \alpha \frac{R}{m} + P_1 * (1-\alpha) \frac{R}{n} \\ \text{s.t.} \quad & \alpha \frac{R}{m} \le \omega_0 \\ & (1-\alpha) \frac{R}{n} \le \omega_1 \\ & 0 \le \alpha \le 1 \end{aligned} \quad (10)$$

The first-order condition (FOC) is $(P_0/m - P_1/n) * R$.

When $FOC > 0$, $1 < \beta < P_0/P_1$.

With $0 \leq \alpha \leq \omega_0 m/R$ or $1 - \omega_1 n/R \leq \alpha \leq \omega_0 m/R$, optimal $\alpha^* = \omega_0 \beta n/R$ and $EU^* = P_0\omega_0 + P_1(R/n - \omega_0\beta)$;

With $0 \leq \alpha \leq 1$ or $1 - \omega_1 n/R \leq \alpha \leq 1$, optimal $\alpha^* = 1$ and $EU^* = P_0 R/m$.

When $FOC < 0$, $\beta > P_0/P_1$.

With $0 \leq \alpha \leq \omega_0 m/R$ or $0 \leq \alpha \leq 1$, optimal $\alpha^* = 0$ and $EU^* = P_1 * R/n$;

With $1 - \omega_1 n/R \leq \alpha \leq \omega_0 m/R$ or $1 - \omega_1 n/R \leq \alpha \leq 1$, optimal $\alpha^* = 1 - \omega_1 n/R$ and $EU^* = P_0(R - n\omega_1)/(n\beta) + \omega_1 P_1$.

2) Meet the demand of all customers with some promotion resources left if all resources focus on one certain promotion channel. Formulate the optimal problem as

$$\begin{array}{ll} \max & P_0 * \omega_0 + P_1 * \omega_1 \\ \text{s.t.} & \alpha \frac{R}{m} \geq \omega_0 \\ & (1-\alpha)\frac{R}{n} \geq \omega_1 \\ & 0 \leq \alpha \leq 1 \end{array} \tag{11}$$

When α meets the constraints above, we solve this inequality constrained optimization problem easily with its optimal solution $EU^* = P_0 * \omega_0 + P_1 * \omega_1$. In reality, situation like this is uncommon as resources are always limited.

3) Resources in hand meet the demand of customers in urgent need of promotion by phone. However, when it comes to putting all resources to the channel of face-to-face promotion, we lack resources to meet the high demand. Describe the optimal problem as:

$$\begin{array}{ll} \max & P_0 * \omega_0 + P_1 * (1-\alpha)\frac{R}{n} \\ \text{s.t.} & \alpha \frac{R}{m} \geq \omega_0 \\ & (1-\alpha)\frac{R}{n} \leq \omega_1 \\ & 0 \leq \alpha \leq 1 \end{array} \tag{12}$$

Formulate the FOC of this inequality constrained optimization problem above as $FOC = -P_1 * R/n$. As $FOC < 0$, with $1 - \omega_1 n/R \leq \alpha \leq 1$, optimal $\alpha^* = 1 - \omega_1 n/R$ and $EU^* = P_0\omega_0 + P_1\omega_1$; with $\omega_0 m/R \leq \alpha \leq 1$, optimal $\alpha^* = \omega_0 m/R$ and $EU^* = \omega_0 P_0 + P_1(R/n - \omega_0\beta)$.

4) Resources in hand meet the demand of customers in urgent need of face-to-face promotion. However, when our optimization solution decides to put all resources to promoting by phone, we are in short of resources. Describe the optimal problem as

$$\begin{array}{ll} \max & P_0 * \alpha * \frac{R}{m} + P_1 * \omega_1 \\ \text{s.t.} & \alpha \frac{R}{m} \leq \omega_0 \\ & (1-\alpha)\frac{R}{n} \geq \omega_1 \\ & 0 \leq \alpha \leq 1 \end{array} \tag{13}$$

The FOC of this inequality constrained optimization problem above is $FOC = P_0 * R/m$. As $FOC > 0$, with $0 \leq \alpha \leq \omega_0 m/R$, optimal $\alpha^* = P_0\omega_0 + P_1\omega_1$; with $0 \leq \alpha \leq 1 - \omega_1 n/R$, optimal $\alpha^* = 1 - \omega_1 n/R$ and $EU^* = P_0(R - n\omega_1)/(n\beta) + P_1\omega_1$.

When $\alpha^* = 1$, we allocate all resources to the stronger strategy of face-to-face promotion, which means the difference between the two channels in promotion cost is small so that this strategy offers a more efficient way to promote a product. In this case, all promoted customers finally buy our new product. In contrast, when $\alpha^* = 0$, all resources in hand are well allocated to the weaker strategy of promotion by phone, indicating that when $\beta > P_0/P_1$, the cost between face-to-face promotion and promotion by phone does show a great difference. The other two optimal cases show that when customer demand is not satisfied with our provided promotion resources, we meet one certain demand in priority and allocate the remaining resources to other customers with different demands.

4. An empirical study for implementing precision marketing strategy

In this section we present a case study from an acquiring agency providing loans to small business merchants. These merchants generally have difficulty to obtain loans via traditional agencies because of their small size, high business risk, lack of collateral, inappropriate management of operations, or their high sensitivity to external factors. Generally, channels for these merchants to obtain funds are mainly traditional banks, agencies providing small amount of loans and other civil channels. Funds needed by small business merchants should be of short-period, frequent, and fast while yet the processing time often takes too long when they apply loans from traditional banks owing to their multi-processing steps. Although small business merchants could get loans from other credit agencies with a shorter processing time and more flexibility, they have to pay a high interest rate or receive approval from collateral, which put much pressure on those who are in need of capital.

The acquiring agency we study appears to meet the large demand market for funds of small business merchants. The acquiring agency provides merchants with acquiring business, who understand their acquiring merchants better. Since the acquiring agency has a point-of-sale (POS) flow of merchants, the business scope is extended by providing an earlier settlement, also as a short-term credit product to those POS acquiring merchants, thus, offering a new way to offer loans based upon credit, mainly the POS flow and their personal information (since most small business merchants are exactly their own legal representatives) instead of collateral. We aim at implementing the proposed precision marketing plan to benefit the acquiring agency when launching new financial products.

With a large scale of loan demand, the differentiation of marketing promotion makes the circumstance even more complex. The most pressing concern is how to recognize various demands in depth. Currently, three major ways related to time period of acquiring settlement are offered:

T+1. The acquiring settlement is completed at a certain time of the next working day;

In time. Settlement is established once transaction occurs;

T+0. Settlement is batched several times in the current working day.

Among these three ways, the interest rate of in time is the highest followed by T+0 and T+1 based on the time the three settlements consume. We aim to explore the features among in time and T+0 merchants as the newly launched product of this acquired agency is somehow similar to in time and T+0 products. T+1 merchants are the target potential customers.

We perform simulation using bootstrap with several types of distribution such as normal and exponential distribution and propose an efficient estimation approach by developing non-linear parametric models to characterize promotion label. We next describe this approach with reference to Kim (2014) [13].

Consider the promotion channel binary, modelled through a Probit model defined by:

$$channel_i = \begin{cases} 1 & if\ X_i\theta + \xi_i \geq 0 \\ 0 & otherwise \end{cases} \tag{14}$$

With the consideration of merchants of high loyalty to T+1 in potential customers, we develop a new way to simulate channel label:

$$channel_i = \begin{cases} 0 & if\ Pr(X_i\theta + \xi_i \leq x) < 0.33 \\ 1 & if\ Pr(X_i\theta + \xi_i \leq x) > 0.67 \\ 2 & otherwise \end{cases} \tag{15}$$

where X_i are characteristics of observable merchants, ξ_i is an error term following certain distribution (e.g., standard normal, uniform or lognormal distribution). Channel valued 0 represents those prefer face-to-face promotion while 1 stands for others who prefer promotion by phone.

Obviously, our training sample are small business merchants with settlement circle selected as in time and T+0 while the test sample are those POS acquiring with T+1. We intend to classify all merchants in the test set according to the similarity of the merchants in the training set. Independent variables X_i are considered as follows:

Approved time. The approved time lasting of an acquiring merchant stands for how long he/she has been an acquired customer of the acquired agency, usually the longer the better, i.e., for merchants with longer approved time, the more products they may experience, which would increase the accuracy of our classification process. Moreover, it may be easier to persuade merchants approved earlier to use a new product as they know our product well and that the new product is beneficial to them;

Gender. A binary variable indicates the gender of our merchant. Consider gender also has an impact on which channel they prefer to be promoted by;

Age. Consider merchants at different life stages have different preferences for the promotion channel of a financial product. Furthermore, age may reflect a common pattern of a time period. For instance, young people today are used to reading messages in WeChat (a multi-function social media mobile application software) while the older generation may prefer making phone calls;

Transaction amount. The transaction amount during the statistical period generally stands for which industry the merchants are in, and how much funds we may offer to them;

Number of transactions. The number of transactions during the statistical period reflects the transaction frequency;

Quality. Quality represents the kind of enterprise that the merchant operates. Small business merchants with different quality have different demands for funds.

Parameters are chosen so as to maintain the same mean and variance of all distributions, which keep the coefficient of variation constant. For each of the distribution outlined above, we simulate the case with low and high variability scenarios, respectively.

Transaction amount less than 1 is excluded in our data set in case for test data or balance inquiry. Among the 537,261 merchants in the data set, 11,342 of them belong to the training sample and the rest the test sample. Merchants without any transaction are also excluded in our data set. Records containing important index of missing value such as certification or register time are excluded as well. Dummy variables are imported to transfer categorical variable into quantitative variables. Variables are all standardized. We then examine the collinearity between variables in our sample after variance analysis and optimize variable combinations.

Set promotion channel as equation 14. The results of accuracy and running time of classification are listed in Table 2 when computed in Intel Core (TM) i5-4200U CPU @ 1.60GHz 2.30GHz and RAM of 4GB with R programming.

Table 2. Evaluation index of different model with equal θ where θ_1 = (1, 1, 1, 1, 1, 1, 1, 1).

Algorithm \ Distribution		Nor	Unif	Exp	Poi	T-test	Weibull	Logit	Lognormal
Running time	D-B	39.49	43.17	44.27	44.10	44.62	41.73	40.73	39.62
	K-NN	83.43	84.14	84.35	84.38	83.20	80.27	79.43	78.75
	SVM	316.80	310.04	278.44	229.26	173.11	274.73	343.42	251.26
Predicted accuracy	D-B	66.22%	63.82%	89.74%	79.07%	80.16%	87.81%	60.86%	75.06%
	K-NN	99.11%	98.98%	99.62%	99.45%	99.72%	99.68%	98.95%	99.33%
	SVM	67.71%	64.99%	94.87%	88.53%	87.89%	93.66%	61.07%	76.56%

As the data shown in Table 2 with an equal weight, by sacrificing sensitivity, the running time shortened significantly based on distance algorithm. When studying each method individually, the distance-based algorithm (D-B) shows an average elapsed time of 42.22 with a range from 39.49 to 44.62 and a variance of 4.44; K-NN 82.24 with a range from 78.75 to 84.38 and a variance of 5.56; SVM 272.13 with a range from 173.11 to 343.42 and a variance of 2948.54. As to the accuracy of each model, K-NN is best fitted with 99.36% on average, followed by SVM 79.41% and distance-based 75.34%.

When the weight in the linear combination of equation 14 varies, different results were obtained. When setting an unequal weight, for instance, θ (5, 5, 10, 20, 20, 1, 1, 1), the running time of the three algorithms remained unchanged on average while the sensitivity of accuracy of both distance-based method and SVM seem to increase significantly. When using the distance-based algorithm, we obtain an average elapsed time of 40.30 ranging from 39.22 to 41.09 with a variance of 0.44 compared to K-NN with an average elapsed time of 83.45 ranging from 80.78 to 84.77 and a variance of 1.47 and SVM, an average elapsed time of 369.49 ranging from 361.69 to 387.24 with variance of 69.35. Moreover, the results show an around 10% increment in classification accuracy of both distance-based method and SVM. Table 3 shows our results in detail.

Table 3. Evaluation index of different model with unequal θ where θ_2 = (5, 5, 10, 20, 20, 1, 1, 1).

Algorithm \ Distribution		Nor	Unif	Exp	Poi	T-test	Weibull	Logit	Lognormal
Running time	D-B	41.09	40.76	41.05	39.22	40.09	39.64	40.21	40.31
	K-NN	83.89	83.11	84.08	83.95	84.77	80.78	84.00	83.05
	SVM	365.51	364.34	370.81	374.16	387.24	362.89	361.69	369.24
Predicted accuracy	D-B	87.86%	87.97%	89.98%	90.02%	90.19%	90.38%	85.01%	76.22%
	K-NN	99.40%	99.55%	99.51%	99.62%	99.62%	99.55%	99.53%	99.32%
	SVM	92.39%	92.09%	96.75%	94.80%	95.49%	96.99%	88.24%	79.13%

K-NN kept a much more stable and high performance at all time while distance-based method and SVM showed an increase in predicted accuracy. Among the eight distributions, when simulated with a T-test distributed error term, both distance-based and SVM algorithms seem to show weaker prediction results than the other four distributions although all accuracy results have reached beyond 85%.

Table 4 shows much higher accuracy when the weight is set as θ = (5, 50, 20, 20, 20, 1, 10, 1). As the experiment was conducted with different parameters in our channel model, the elapsed time of SVM varies greatly while that of the distance-based algorithm and K-NN remain rather stable.

Table 4. Evaluation index of different model with unequal θ where θ_3 = (5, 50, 20, 20, 20, 1, 10, 1).

Algorithm \ Distribution		Nor	Unif	Exp	Poi	T-test	Weibull	Logit	Lognormal
Running time	D-B	41.00	39.53	39.43	39.75	39.44	40.32	40.27	39.97
	K-NN	84.23	82.91	81.40	85.32	85.11	81.72	81.85	81.33
	SVM	75.78	81.68	71.17	59.41	53.46	67.60	76.91	63.35
Predicted accuracy	D-B	97.18%	97.20%	97.45%	98.24%	98.10%	97.73%	96.76%	90.99%
	K-NN	99.66%	99.84%	99.84%	99.77%	99.80%	99.72%	99.74%	99.24%
	SVM	97.87%	97.79%	98.14%	98.51%	98.42%	98.18%	97.36%	91.08%

Our results show that the distance-based algorithm always performs the fastest among the three methods although sensitivity of accuracy in Table 4 shows little difference. Accuracy has improved with parameters when the promotion channel varies. With a higher weight on gender, the models are much fitted. To our surprise, SVM requires less time to yield more accurate results. The mean value of elapsed time of distance-based algorithm is 39.96 ranging from 39.43 to 41.00 with a rather low variance of 0.30; K-NN shows a mean elapsed time of 82.98 ranging from 81.33 to 85.32 with a variance of 2.81; while SVM shows a mean elapsed time of 68.67 ranging from 53.46 to 81.68 with a variance of 91.50, a much lower value when compared to itself under equal weight θ.

Among the eight distributions mentioned above, K-NN performed stable enough that it did not show much difference in both running time and prediction accuracy while the running time of our distance-based approach remained almost unchanged but with decreased prediction accuracy in T-test. The difference found in SVM seems more likely to be random as we did not find any regular pattern.

The acquiring agency provided two types of promotion channel: face-to-face promotion and promotion by phone. In fact, promotion resources sometimes do not meet merchants' demand, resulting in four scenarios: i) all resources are allocated properly without any left; ii) all merchants get promoted properly with some resources left; iii) demand from merchants requiring face-to-face promotion is greater than the exact resources while resources of the other channel is adequate, and then the remaining resources of promotion by phone are allocated to merchants who are in need of face-to-face promotion; iv) case 4 is the opposite of case 3 where resources of face-to-face promotion is adequate while promotion by phone is not, we meet the demand of face-to-face promotion firstly and allocate the remaining resources to promotion by phone.

When on-line learning was combined with machine learning algorithms, we searched for a stable predicted accuracy in each type of algorithm and the results are presented in Table 5 and Table 6. Two types of learning rate are presented: one for equal rate and the other for exponential rate, which is determined by the number of observations in each round. The distance-based algorithm found its stable parameters of P_0 and P_1 quickly by round 2 in a total of 10 rounds while K-NN reached its stable predicted accuracy after round 5 in a total of 10 rounds. The distance-based approach is the fastest algorithm to reach its stable prediction, usually in the second round. As it always takes a lot of other resources such as time and capital to get feedback after promotion in each round, the longer it takes for us to reach potential customers, the more likely they might be promoted by other companies with similar products because of the competitive environment. The distance-based algorithm works well to help us deal with this situation and performs precision marketing quickly.

Table 5. Stable status of on-line learning under equal learning rate.

	Round	P_0	P_1	Round	P_0	P_1
D-B	2/10	92.85%	90.72%	2/50	93.22%	93.15%
K-NN	5/10	97.12%	98.72%	21/50	97.47%	98.77%
SVM	6/10	98.33%	98.93%	20/50	98.10%	98.55%

Table 6. Stable status of on-line learning under exponential learning rate.

	Round	P_0	P_1	Round	P_0	P_1
D-B	2/4	93.07%	90.82%	2/6	93.48%	92.14%
K-NN	4/4	97.72%	99.13%	6/6	97.54%	99.05%
SVM	4/4	98.56%	99.02%	6/6	98.45%	98.93%

Later on, we discuss the results of our classification algorithm together with an optimal allocation plan compared with the market planning that the acquiring agencies use currently. The latter is considered to be the benchmark. Estimate the promotion preference of merchants above as an input of the optimal allocation proposal, i.e., once the preference towards promotion channel of customers is recognized, we get a predicted table containing parameter P_0 and P_1 similar to Table 1. Merchants with no interest are excluded in this part to simplify the case. Solve the optimized problem of equation 9. Experiments were conducted to simulate EU^*. Parameters are set as: $R = 1000000$, $m = 200$, $n \in (199,100)$, $\omega_0 \in (4000,6000)$, $\omega_1 \in (4000,12000)$. P_1 is generated with $\theta = (5,5,10,20,20,1,1,1)$ while $P_0 = 1$ in this case. Simulation was conducted with 1000 random combinations of these parameters, covering all analysed scenarios. The ratio of the marketing strategy adopted by the acquired agency currently followed by resource allocation is $50/50$, i.e., among all resources R, 50% are allocated to face-to-face promotion while the remaining 50% to promotion by phone.

The simulation process has four steps: i) generate merchants' channel label; ii) generate merchants' preference towards promotion channel; iii) solve optimization problem of resource allocation; and iv) iteration back to step 1 until the simulation time is larger than 1000. We obtain P_0, P_1 by different classification algorithms in step 2 then with a random combination of parameters related to resource allocation, we obtain the optimal EU^* after step 3. Table 7 shows the results of average EU^* of mixed arrangements. The results show that disregard the classification algorithm used for the prediction, the optimal resource allocation always generates a larger number of expected users than that of the random allocation currently used by the acquired agency despite the small difference among the three methods. Comparisons were also made. As we set the bundle of distance-based classification and the random allocation as the benchmark (which performed the weakest), K-NN with optimal resource allocation plan improved most with up to 22.42% improvement. Our results show that the optimal resource allocation help to increase the number of expected users significantly.

Table 7. EU^* under different marketing strategies.

	Optimal resource allocation		Random allocation in use	
	EU^*	Improvement	EU^*	Improvement
Distance-based	6638	21.60%	5459	Benchmark
K-NN	6683	22.42%	5835	6.89%
SVM	6654	21.89%	5689	4.21%

5. Conclusions

In this research, we propose a general framework for precision marketing aiming at promoting new products. A customer prediction model including distance-based algorithm, K-NN and SVM towards promotion channel was presented to forecast the demand of different promotion channels, and later combined with on-line learning to fit a dynamic situation. The classification models proposed was able to extract the characteristics of the relationship between customers and historical products. Later on, the resource allocation problem was formulated to find the optimal solution through classification results accurately.

Our empirical study shows that the customer segmentation based on center distance is accurate and help companies to identify potential customers by minimizing the corresponding errors in marketing decisions. Based on our findings, enterprises could make precise marketing strategies for different customer categories. In addition, our case study shows that our precision marketing strategy is efficient and help enterprises in planning their marketing promotion. The classification results show that with a fairly large sample of 525,919 merchants for classification, the general distance-based algorithm is most efficient in generating available results while K-NN performed most stable in terms of accuracy but requires twice as much time as that of the distance-based method. SVM is slightly more accurate in its prediction than the distance-based method, but requires a long time to process owing to obtaining its convex optimization results. We also recognize customers with little or no interest in new products through on-line learning. Among the three methods, distance-based algorithm is most efficient as it costs least to reach its stable predicted accuracy, usually in the second round. Later on, we take market demand and the resources supplied into consideration as the demand of customers is always uncertain and our promotion resources in hand are limited. To solve the optimization problem, we present an allocation plan covering four scenarios by taking the uncertain merchant demand into account: i) promotion resources are abundant enough to be allocated to all merchants; ii) resources remain inadequate when allocated to any one type of merchants; iii) resources are adjusted to be only enough to be allocated to merchants who are in favour of face-to-face promotion; and iv) resources are adjusted to be only enough to be allocated to merchants who prefer promotion by phone. 1000 simulations were performed and finally we obtained an average of the improved results. When our classification results were combined with the allocation strategies, we obtain an increase in expected users by up to 20%, among the three methods, with the K-NN classification results outperforming others with an 22.42% increase, suggesting the recognition of which type of channel merchants prefer is vital when doing precision marketing.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) [Grant No: 71971168].

References

[1] Rust R, Zeithaml V and Lemon K 2004 Customer-centered brand management *Harvard Business Review* **82(9)** 110-120
[2] Gartner D, Kolisch R, Neill D B and Padman R 2015 Machine learning approaches for early drg classication and resource allocation *INFORMS Journal on Computing* **27(4)** 718-734
[3] Huang D and Luo L 2016 Consumer preference elicitation of complex products using fuzzy

support vector machine active learning *Marketing Science* **35(3)** 445-464
[4] Galindo J and Tamayo P 2000 Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications *Computational Economics* **15(1)** 107-143
[5] Zhang J L and Härdle W K 2010 The bayesian additive classication tree applied to credit risk modelling *Computational Statistics & Data Analysis* **54(5)** 1197-1205
[6] Turney P D 2002 Thumbs up or thumbs down? Semantic orientation applied to unsupervised classication of reviews *Proc. 40th annual meeting on association for computational linguistics* (Association for Computational Linguistics) pp 417-424
[7] Cui G, Wong M L and Lui H 2006 Machine learning for direct marketing response models: Bayesian networks with evolutionary programming Management Science **52(4)** 597-612
[8] Lin Z, Chen G, Bai X, Lv H, Yin W and Dong J 2009 Customer clustering using semi-supervised geographic information *IEEE/INFORMS International Conference on Service Operations, Logistics and Informatics* pp 465-470
[9] Kim Y, Street W N, Russell G J and Menczer F 2005 Customer targeting: A neural network approach guided by genetic algorithms *Management Science* **51(2)** 264-276
[10] Marcus C 1998 A practical yet meaningful approach to customer segmentation *Journal of consumer marketing* **15(5)** 494-504
[11] Hochbaum D S, Moreno-Centeno E, Yelland P and Catena R A 2011 Rating customers according to their promptness to adopt new products *Operations research* **59(5)** 1171-1183
[12] Hinton G E, Osindero S and The Y 2006 A fast learning algorithm for deep belief nets *Neural computation* **18(7)** 1527-1554
[13] Kim S, Chan C W, Olivares M and Escobar G 2014 ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes *Management Science* **61(1)** 19-38
[14] Wang G, Li J, Hopp W J, Fazzalari F and Bolling S F 2017 Using Patient-Specific Quality Information to Unlock Hidden Health Care Capabilities *Ross School of Business Paper* No. 1291
[15] Ban G and Rudin C 2019 The big data newsvendor: Practical insights from machine learning *Operations Research* **67(1)** 90-108
[16] Fischer M, Albers S, Wagner N and Frie M 2011 Practice prize winnerdynamic marketing budget allocation across countries, products, and marketing activities *Marketing Science* **30(4)** 568-585
[17] Hanssens D M, Parsons L J and Schultz R L 2003 Market response models: Econometric and time series analysis vol 12 (Springer Science & Business Media)
[18] Zinkevich M 2003 Online convex programming and generalized infnitesimal gradient ascent *Proc. 20th International Conference on Machine Learning* pp 928-936
[19] James G, Witten D, Hastie T and Tibshirani R 2013 *An introduction to statistical learning* vol 112 (Springer)
[20] Aslay C, Lu W, Bonchi F, Goyal A and Lakshmanan L V 2015 Viral marketing meets social advertising: Ad allocation with minimum regret *Proc. VLDB Endowment* **8(7)** 814-825
[21] Metzen J H 2016 Minimum regret search for single-and multi-task optimization *Proc. The 33rd International Conference on Machine Learning* **48** 192-200
[22] Eduardo C 2017 A minimum expected regret model for the shortest path problem with solution-dependent probability distributions *Computers & Operations Research* **77** 11-19