Optimizing Marketing Campaign Targeting Using Uncertainty-Based Predictive Modelling

Björn Rafn Gunnarsson, Seppe vanden Broucke and Jochen De Weerd

Department of Decision Sciences and Information Management

KU Leuven, Naamsestraat 69, B-3000 Leuven

bjornrafn.gunnarsson@kuleuven.be; seppe.vandenbroucke@kuleuven.be; jochen.deweerd@kuleuven.be

Abstract—Over the past decades, decision makers have increasingly relied on analytical methods in order to increase revenues and reduce costs. This is also the case for marketing, where predictive models have primarily been used to identify potential customers and measure the effects of marketing campaigns. In this paper, we introduce a methodology to optimally distribute a constrained marketing budget over a group of targets for which historical behavioral information is available, though where it is deemed infeasible to set up a trial campaign involving a control and test group as is the typical approach described by “net lift” modelling. Instead, so called “swing clients” are identified based on a notion of uncertainty following from any classification technique which can be trained over the historical data, which makes that our approach is easily applicable in most marketing environments. We report on the feasibility of our approach by presenting a case study which was performed at the largest airport in Belgium.

Keywords: Marketing analytics, Propensity maximization, Market potential estimation, Marketing budget allocation, Market Campaign Targeting.

I. INTRODUCTION

Marketing activities have become an essential component of many businesses in order to ensure their organic growth. In the last decade, companies in the U.S. allocated billions of dollars to marketing, which gives a clear indication of the importance of marketing for today’s companies. However, marketing strategies which are directed to random customers often generate huge cost with relatively little return. Hence, it is clear that marketing managers have a considerable responsibility to allocate these constrained marketing funds in an optimal way, e.g. between different customer segments, product groups, or geographical areas, in order to generate appropriate returns for their organization. Over the past decades, decision makers have increasingly relied on analytical methods in order to increase revenues and reduce costs. This is also the case for marketing, where predictive models have been primarily used to identify potential customers and measure the effects of marketing campaigns. Analytical methods have also been developed to optimally distribute constrained marketing budgets. Some pioneering examples of research on how to optimally distribute a marketing budget across geographical areas include [1] and [2]. However, the general focus of marketing research shifted away from this topic in the 1980s towards empirical estimation aspects of marketing problems [3]–[8].

In this paper, we introduce a methodology to optimally distribute a constrained marketing budget across target groups. The proposed solution is based on maximizing the potential of a marketing campaign by estimating the overall likelihood that it can obtain new customers. Our approach was developed and applied in the real-life setting of a large international airport in Belgium, namely Brussels Airport, to optimally distribute a constrained marketing budget across geographical areas when promoting the airports services within Belgium. We will utilize this application context in what follows to highlight the feasibility of our approach in practice.

The methodology for distributing a constrained marketing budget suggested here draws inspiration from propensity scoring methods originally suggested by [9]. These methods have been applied in a number of different fields since their introduction more than three decades ago, including marketing where they have traditionally been used to rank a list of individuals to contact, from the most likely to generate additional revenue to least likely to do so. This is done so that individuals that are thought to provide the best return on investment will be contacted first.

More recently, another class of models called “net lift” models have been developed that aim at maximizing the true impact of a marketing campaign by focusing on “swing clients”, i.e. customers which will not buy products or react in a positive manner spontaneously (cases which would be ranked highly by propensity scoring methods but would still incur a cost when targeted as no additional effort needs to be allocated to this group), but need to be motivated (often including a form of incentive) to take action. Typically, such models are constructed by splitting the campaign population into a test group, which receives the marketing campaign, and a control group, which does not receive the campaign. Net lift modelling then aims at finding a model such that the difference between the test group purchase rate and the control group purchase rate is maximized in such a way that identifies the swing clients [4]–[6], [10]. A drawback, however, of this approach is that information on the purchase rate of the two groups is generally not readily available from historical data sources, making this setup infeasible or costly in most practical settings.

The motivation of this paper is therefore to construct a method for estimating market potential similar to net lift modeling using limited historical data. To this end, we propose a novel framework to utilize standard well-known classification

techniques in a setting where propensity maximization is used to estimate market potential when only historical factual data is available, based on the notion of uncertainty of a probabilistic classifier. This underlines the main contribution of this paper: A straightforward but powerful technique is proposed for estimating market potential, which can be easily constructed using typical historical data sources. Furthermore, we illustrate the applicability of this approach by reporting on a use case where a constrained marketing budget was distributed between geographical areas within Belgium, where the methodology described by our proposed framework was combined with other public data sources to form an ensemble approach for estimating market potential within a given area.

The remainder of this paper is structured as follows. Section II provides a detailed description of uncertainty based approach proposed in this paper. Section III details a use case where this method was combined with other data sources to construct an ensemble approach for distributing a constrained marketing budget between geographical areas. Finally, some concluding remarks will be given in the last section of the paper.

II. AN UNCERTAINTY-BASED APPROACH FOR ESTIMATING MARKETING POTENTIAL

Our approach was developed and applied in the real-life setting of a large international airport in Belgium, where the goal was to allocate a marketing budget over different municipalities (geographic areas) in Belgium and where demographic characteristics of the residents were available, as well as whether they had used the airport as their main transportation provider or, alternatively, had chosen a competitor. We will utilize this setting as a running example to describe the setup of our approach.

Recalling the remarks provided in the introductory section, it is important to highlight that the goal of an analytical approach in this context is to maximize the potential of the marketing campaign. A direct propensity-based approach would be able to rank residents (and broader geographic areas) in terms of how likely they are to opt for the firm’s services compared to a competitor based on their characteristics. Nevertheless, high-ranking cases need not receive a high degree of budget allocation, since these cases have no need to be re-targeted again—forming the group of “sure things” as typically described in the area of net lift modeling. Similarly, lower-ranking cases—the “lost causes”—are as convinced to utilize a competitor’s offer and are hence very unlikely to be persuaded, making budget allocation to this group a sub-optimal outcome as well. “Net lift” modeling approaches hence try to identify the group of persuadable instances, but often come with stringent data requirements in the form of a control and test group which is often unavailable or difficult to obtain in practical settings [5], [6].

Hence the uncertainty approach suggested here utilizes a binary classifier in order to estimate how likely instances are to respond positively to products or services. In order to depict the development of the classifier, let $x = (x_1, x_2, \dots, x_m) \in$

Figure 1: An area of uncertainty is defined over the probability distribution of a constructed binary classifier

$\mathbb{R}^m$ be an m -dimensional vector describing characteristics of an instance x and let $y \in \{-1, +1\}$ be a binary variable that distinguishes between instances that bought products or utilized services from the company ($y = +1$) and instances who did not ($y = -1$). A classifier $\mathcal{L} : \mathbb{R}^m \mapsto [0, 1]$ is constructed that estimates the probability $p(+|x)$ that a purchase will be made for resident x , where $p(+|x)$ is a shorthand form of $p(y = +1|x)$. A large number of such classification models have been suggested in the literature including traditional statistical methods (e.g. logistic regression), non-parametric statistical models (e.g. decision trees) and neural networks (for an extensive overview of classification models see e.g. [6] or [11]).

Any binary classifier could in theory be used in the framework suggested here for estimating the likelihood of an instance using the offered services or products. When developing our approach, a decision tree algorithm was used for estimating the likelihood of using the firm’s services ($p(+|x)$) given an instance (x). These algorithms are among the most well know and widely used of all machine learning algorithms and come up with a tree like structure representing patterns in an underlying data set by recursively partitioning it based on its attributes into subgroups that are as homogeneous as possible in terms of the variable to be predicted [6], [12]–[15].

In our particular setting, this classifier was constructed using data obtained from a survey carried out among individuals living in Belgium that had travelled in the past year. The data contained demographic information on residents and information on what airport they had used in the past year. The latter was transformed into a binary target variable indicating if residents had used Brussels Airport when travelling ($y = +1$) or the services of one of its competitors ($y = -1$). The resulting data set was used to construct a classification model for each geographical area that estimates the probability of individuals using Brussels Airport based on their demographic information. The obtained probability estimates $\{\mathcal{L}(x)\}$ can

Figure 2: Map showing the proportion of residents in each area based on the uncertainty-based approach. Darker colored areas contain more clients that can still be captured by marketing efforts. Lighter colored areas contain clients that are already loyal customers or cannot be convinced to switch

be used to construct a probability distribution in each area which shows the density of the overall likelihood of residents using Brussels airport in that area. An example of the obtained probability distribution for one area in Belgium is given in Figure 1.

Recall our aim to distinguish between three types of potential customers. First, there are individuals that would never buy the product or use the provided services in question, irrespective of them being exposed to a marketing campaign. As noted above, this group of individuals can be considered “lost causes” and targeting them would not be sensible since they won’t buy the product anyway. Second, there are individuals that can be considered “sure things” and targeting them will cause a loss in profit since they will always buy the product, therefore offering them a marketing incentive (e.g. a discount

or promotional offer) will reduce the profit margin of the company. The third category of potential customers is the “swing clients”. This is a group of individuals that will not buy the product spontaneously, but still exhibit potential to be captured or acquired by the marketing effort. Because of this, a marketing campaign can be especially effective for these individuals and focusing marketing efforts on this group of potential customers will maximize the impact of the marketing campaign [6].

To take this into account, the uncertainty approach proposed here defines an area of interest of the probability distribution of the constructed classifier. This is an area of the distribution which only includes the instances in each area that are undecided on whether to buy a product or utilize services from the company that is distributing its marketing budget (swing

clients). In other words, it includes those instances for which it is still difficult to predict their usage pattern with a high degree of certainty, encompassing the “lost causes” and “sure things” we do not aim to capture in the campaign. This area can be determined based on expert knowledge or statistically. For instance, in Figure 1, the area of interest is defined to be in the interval between 0.3 and 0.8, targeting 45% of the residents, which can be considered swing clients which would form the focus group of marketing efforts.

Alternatively, the range of uncertainty of the constructed classifier can also be derived statistically. To do so, we assume the probability distribution of the classifier is well calibrated [16]. Well calibrated classifiers are probabilistic classifiers for which the probabilistic outputs can be directly interpreted as a confidence level. That is, a well calibrated binary classifier should classify instances such that, for instances for which a probability close to $p(+|x_i)$ is provided, approximately the same ratio of instances found in the population actually belong to the positive class. Even although many binary classifiers are not well calibrated (including decision trees), various techniques exist that transform the probability distribution to a well calibrated distribution. Next, a beta distribution, which forms an appropriate distribution to describe the output of a probabilistic binary classifier [17] with probability density function (PDF) $\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ (with B a normalization constant to ensure the cumulative probabilities sum to 1) is fitted over the probability distribution of the classifier $\mathcal{L}$, estimating α and β (for instance using the well known method of moments technique [18]). Next, lower and upper bounds for the uncertainty range can be defined by solving the following optimization problem:

$$\begin{aligned} \min \quad & u - l \\ \text{with} \quad & I_u(\hat{\alpha}, \hat{\beta}) - I_l(\hat{\alpha}, \hat{\beta}) \geq M \\ & l \in [0, 0.5] \\ & u \in [0.5, 1] \end{aligned}$$

With $\hat{\alpha}$ and $\hat{\beta}$ the fitted parameters for the beta distribution fitted over the probability distribution of the classifier, $I_x(\alpha, \beta)$ the cumulative distribution function (CDF) for the beta distribution evaluated at x , and l and u the lower and upper bound respectively. These define the minimum region of uncertainty so that the probability mass in the region exceeds a given desired uncertainty mass value M . In our experiments, we found the value of 0.2 to be a good threshold, though it can be adjusted based on domain expertise and will effectively determine how much customers will be considered as being part of the uncertainty range. Bounds for l and u are set below and above 0.5 under the assumption of a well calibrated classifier as outlined above. Note that solving this optimization problem is hard given the non-linear cumulative distribution function and can be solved using appropriate non-linear optimization techniques such as the interior point method [19]. Alternatively, optimal values for l and u can also easily be empirically estimated using a grid search, which, given the problem statement, is easy to compute.

Figure 3: Map showing whether areas are close to Brussels Airport or a competitor airport, or fall in between

Finally, note that here too, the region of uncertainty over the classifier’s output $[I_l(\hat{\alpha}, \hat{\beta}), I_u(\hat{\alpha}, \hat{\beta})]$ captures a group of instances which will be included in the marketing target group. That is, an instance x is included if $I_l(\hat{\alpha}, \hat{\beta}) \leq \mathcal{L}(x) \leq I_u(\hat{\alpha}, \hat{\beta})$.

III. USE CASE: ENSEMBLE APPROACH

The uncertainty-based approach discussed in the previous section was used to estimate the market potential of Brussels Airport in different areas (municipalities) within Belgium. This approach measures the market potential within areas by estimating the proportion of swing clients within each area as follows. The uncertainty approach described above was applied to identify the swing clients in all geographical areas included in the campaign to be set up, aggregating results to a proportion describing the amount of swing clients for each geographic area. The proportion of swing clients in all municipalities in Belgium is shown in Figure 2. As can be seen from the figure, the approach is able to make a clear distinction between the areas with more and less swing clients. A large shoe-shaped area is identified where the marketing potential is estimated to be quite low (0-10%). This is an area which already has a high number of loyal customers that will continue to use the airport’s services irrespective of a marketing campaign and should therefore not be the focus of the airport’s marketing efforts. Also, a number of municipalities are identified that are estimated to have a high proportion of swing clients and therefore a high potential for a marketing campaign ($\geq 60\%$). These are areas which should be the focus of the airport’s marketing efforts.

The idea is now to distribute the constrained marketing budget according to the proportions provided by the uncertainty-based approach. However, the size of the population in each area is also an important consideration in this setting, as

Figure 4: Overview of the final ensemble model constructed to guide the budget allocation of the marketing campaign, combining the uncertainty-based approach with other ratios

very populated areas will intuitively contain a higher number of swing clients than areas with a very small population, which is not taken into account by the aggregated proportions. Therefore, this should be taken into account as well so that more marketing efforts are distributed to areas with larger populations and less efforts are distributed to areas with smaller populations.

Additionally, Brussels Airport and other competing airports provide their services in specific geographical locations. Therefore, the proximity of an area to an airport is an important indicator of residents in that area using that airport. To take this into account the distance from the center point of each postal code in Belgium to all competing airports in Belgium was computed and used to compute a distance ratio that indicates if an area is close to Brussels airport, a competing airport, or falls between the airports in Belgium as shown in Figure 3. This ratio was then turned into a normalized ratio where a high value indicates that the area is close to a competing airport and a low score indicates that the area is close to Brussels airport.

Finally, an income-based ratio was also constructed for inclusion in the complete model. Although popular, travelling and consequently the services of airports are not essential goods for people in general. Therefore, the average income in each area is an important indicator of residents using the services of the airport. In order to take this into account, the fraction of the income in each area compared to average income in Belgium (using publicly available governmental information) was computed.

All the indicators discussed in this section were combined

with the outcome of the uncertainty-based approach, as shown in Figure 4, to form a combined ensemble model for marketing budget allocation. The different approaches were combined using multiplication and the subsequent result was transformed by a softmax function:

$$\sigma(z)_j = \frac{\exp(z_j)}{\sum_{k=1}^K \exp(z_k)}$$

for $j = 1, \dots, K$. As shown the function applies the standard exponential function to each element z_j of a k dimensional input vector z and normalizes these values by dividing by the sum of all these exponentials [20]. In our case the input vector consists of the outcomes of multiplying the result of the four constructed models for each area. By using the softmax function we obtain scores per area which are bounded between 0 and 1 (and sum to 1) and can therefore be interpreted as the proportion of the marketing budget that should be allocated to each area in Belgium.

This ensemble combines the benefits of a number of different data sources and is therefore able to distribute the fixed marketing budget in a more optimal way than if one would only use each of the different data sources separately. The ensemble aims at maximizing the effect of the campaign by distributing a higher proportion of the marketing budget to more populated areas with a high estimated potential of a marketing campaign and a low proportion of the marketing budget to less populated areas with little estimated market potential.

Figure 5 shows the amount of euros the ensemble would distribute to each area in Belgium based on a hypothetical

Figure 5: Map showing an overview of marketing budget allocation over municipalities of Belgium as determined by the final ensemble model

situation where a marketing budget of 250 thousand euros is being distributed. The two areas the ensemble allocates the highest amount towards are Antwerp and Liege. The areas are among the most highly populated areas in Belgium and both are estimated to have a considerable potential for a marketing campaign. Areas that fall in the shoe like area inhabited by loyal customers of Brussels airport as discussed above get a low amount of the total marketing budget relative to their population size.

To the best of our knowledge, the framework suggested here is the first approach that utilizes a classification model to estimate market potential based on the concept of uncertainty to represent potential to allocate a constrained marketing budget, underlying the novelty of this method.

IV. CONCLUSION

Modern businesses allocate vast amounts of resources to marketing efforts in order to amplify their growth. Marketing strategies where random customers are targeted can however often generate huge costs without adequate returns. Analytical method can be effectively used to aid marketing managers in their decision making in order to increase revenue and reduce costs. This paper suggests a novel and general framework for estimating market potential within geographical areas which consequently can be used to optimally distribute a fixed marketing budget. Previous methods suggested in the literature which aim to maximize the true impact of marketing campaigns utilize information on the purchasing rates of a test group, which receive a marketing campaign, and a control group, which do not receive the campaign. However, this information is often not readily available from historical data

sources and needs to be collected before the methods can be applied.

Therefore, a straightforward but powerful uncertainty-based approach for measuring marketing potential was suggested in this paper, where only historical factual data is available. The method utilizes a binary classifier which estimates the likelihood of residents in an area of buying the product or service which is to be promoted using a marketing campaign. These estimates are then used to identify areas with a high proportion of residents that can be considered to be swing clients. That is, the group of customers which will not buy products or services spontaneously but can still be incentivized to do so and should therefore be the focus of the marketing campaign.

The applicability of this method was illustrated through a use case where a constrained marketing budget was distributed between geographical areas within Belgium, where the uncertainty-approach suggested here was combined with other public data sources to form an ensemble approach to allocate the constrained marketing budget within Belgium.

REFERENCES

- [1] G. L. Urban, "Allocating ad budgets geographically," *Journal of Advertising Research*, vol. 15, no. 6, pp. 7–16, 1975.
- [2] D. M. Holthausen Jr and G. Assmus, "Advertising budget allocation under uncertainty," *Management Science*, vol. 28, no. 5, pp. 487–499, 1982.
- [3] S. Gupta and T. Steenburgh, "Allocating marketing resources," *Marketing Mix Decisions: New Perspectives and Practices*, Roger A. Kerin and Rob O'Regan, eds., American Marketing Association, Chicago, IL, pp. 90–105, 2008.
- [4] P. Rzepakowski and S. Jaroszewicz, "Uplift modeling in direct marketing," *Journal of Telecommunications and Information Technology*, pp. 43–50, 2012.
- [5] A. Kumar and R. Kumar, "Uplift modeling: Predicting incremental gains," 2018.
- [6] B. Baesens, *Analytics in a big data world: The essential guide to data science and its applications*. John Wiley & Sons, 2014.
- [7] B. Leventhal and S. Langdell, "Adding value to business applications with embedded advanced analytics," *Journal of Marketing Analytics*, vol. 1, no. 2, pp. 64–70, 2013.
- [8] M. K. Mantrala and V. K. Kanuri, "14. marketing optimization methods," *Handbook of Marketing Analytics: Methods and Applications in Marketing Management, Public Policy, and Litigation Support*, p. 324, 2018.
- [9] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.
- [10] D. B. Rubin and R. P. Waterman, "Estimating the causal effects of marketing interventions using propensity score methodology," *Statistical Science*, pp. 206–222, 2006.
- [11] M. Kuhn and K. Johnson, *Applied predictive modeling*. Springer, 2013, vol. 26.
- [12] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," *Journal of the operational research society*, vol. 54, no. 6, pp. 627–635, 2003.
- [13] S. L. Salzberg, "C4. 5: Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993," *Machine Learning*, vol. 16, no. 3, pp. 235–240, 1994.
- [14] S. Sharma, J. Agrawal, and S. Sharma, "Classification through machine learning technique: C4. 5 algorithm based on various entropies," *International Journal of Computer Applications*, vol. 82, no. 16, 2013.
- [15] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, "A comparative study of decision tree id3 and c4. 5," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 2, 2014.
- [16] M. Kull, T. M. Silva Filho, P. Flach *et al.*, "Beyond sigmoids: How to obtain well-calibrated probabilities from binary classifiers with beta calibration," *Electronic Journal of Statistics*, vol. 11, no. 2, pp. 5052–5080, 2017.
- [17] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.
- [18] D. C. Montgomery, G. C. Runger, and N. F. Hubele, *Engineering statistics*. John Wiley & Sons, 2009.
- [19] N. Karmarkar, "A new polynomial-time algorithm for linear programming," *Combinatorica*, vol. 4, no. 4, pp. 373–395, Dec 1984.
- [20] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.