



A VNS approach for book marketing campaigns generated with quasi-bicliques probabilities

Thays A. Oliveira^{a,b,c, 1}, Vitor N. Coelho^{c, 2},
Helena Ramalhinho^d, Marcone J. F. Souza^e,
Bruno N. Coelho^c, Daniel C. Rezende^f and Igor M. Coelho^{c,g}.

^a*Business and Economics Graduate Program at Universidade Federal de Lavras*

^b*Information and Communication Technologies Graduate Program at Universitat Pompeu Fabra, Barcelona, Spain*

^c*Grupo da Causa Humana, Ouro Preto, Brazil*

^d*Dpt. of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain*

^e*Dpt. of Computer Science at Universidade Federal de Ouro Preto*

^f*Dpt. of Business and Economics at Universidade Federal de Lavras*

^g*Dpt. of Computer Science at Universidade Estadual do Rio de Janeiro*

Abstract

This paper focuses on Book Marketing Campaigns, where the benefit of offering each book is calculated based on a bipartite graph (biclique). A quasi Biclique problem is assessed for obtaining the probabilities of success of a given client buy a given book, considering it had received another book as free offer. The remaining optimization decision problem can be solved following the Targeted Offers Problem in Direct Marketing Campaigns. The main objective is to maximize the feedback of customers purchases, offering books to the set of customers with the highest probability of buying others ones from its biclique and, at the same time, minimizing campaign operational costs. Given the combinatorial nature of the problem and the large volume of data, which can involve real cases with up to one million customers, metaheuristics procedures have been used as an efficient way for solving it. Here, a hybrid trajectory search based algorithm, namely GGVNS, which combines the

Greedy Randomized Adaptive Search Procedures and General Variable Neighborhood Search, is used. The strategy for generating the quasi Biclique problem is described and a new instance generator for the TOPDMC is introduced. Computational results regarding the GGVNS algorithm shows it is able to find useful and profitable sets of clients.

Keyword: Books marketing, Campaigns, Targeted offers problem, Quasi-Biclique, General Variable Neighborhood Search, Operational Research

1 Introduction

Currently the field of literature has undergone several transformations, in particular, with the advent of new technologies and mobile devices, reading is becoming increasingly widespread. Big-data datasets regarding the purchase of books will keep growing. Edges connecting clients and books can be used to represent links between them. Mining information from those bipartite graphs (biclique) is the task introduced and discussed in this paper.

The problem of finding suitable books to be offered to a given client can fit the scope of the Targeted Offer Problem in Direct Marketing Campaign (TOPDMC) [5], that seeks to select the most profitable set of customers for offering products in Direct Marketing campaigns. Generally speaking, the main goal is to find an appropriate set of clients to receive a book offer, maximizing campaign profits while respecting operational requirements.

We consider the real case where a first book is offered as gift [3], expecting that this fact will reinforce and strength biclique connections (as can be seen in Figure 1). Considering this new offer, the probability of that client buying other recommended books, from its quasi-biclique, is measured. Examples of free samples and consumers purchase have been studied in different sectors [3]. Lammers [3] verified that sampling significantly increased the immediate sales of chocolates, however, most part of the clients purchased chocolate varieties other than the variety sampled. Cosmetic companies have been doing free sampling [1] and still appear to be open for new strategies based on Artificial Intelligence (AI) tools.

The sum of the probabilities of a given client i buying a book, considering it received book j as an offer, are used for creating an instance of the TOPDMC. The latter is solved using a previous designed Variable Neighborhood Search

¹ Email: thaysoliveira7@gmail.com or thaysaparecida.deoliveira01@estudiant.upf.edu

² Email: vncoelho@gmail.com

algorithm. Figure 1 represents a single biclique, connecting clients and books. Books may have already been read or might be offered or recommended during the marketing campaign. Thus, each client i might had read some books, let consider a set of books $B_i = \{b_i^1, \dots, b_i^b, \dots, b_i^m\}$, where m is number of books read by that client. In this case, an edge connects clients i to each read book b_i^b from this set B_i . In the depicted example, the following books were already read, bought or received as gift: $B_1 = \{b_1^3\}$; $B_2 = \{b_2^1, b_2^2\}$; $B_3 = \{b_3^3, b_3^4\}$. As can be noticed: books 1, 2 and 4 still can be offered or recommended to the client 1; book 1 is being offered as a free sample while books 2 and 4 are being recommended. Similar cases are exemplified for clients 2 and 3.

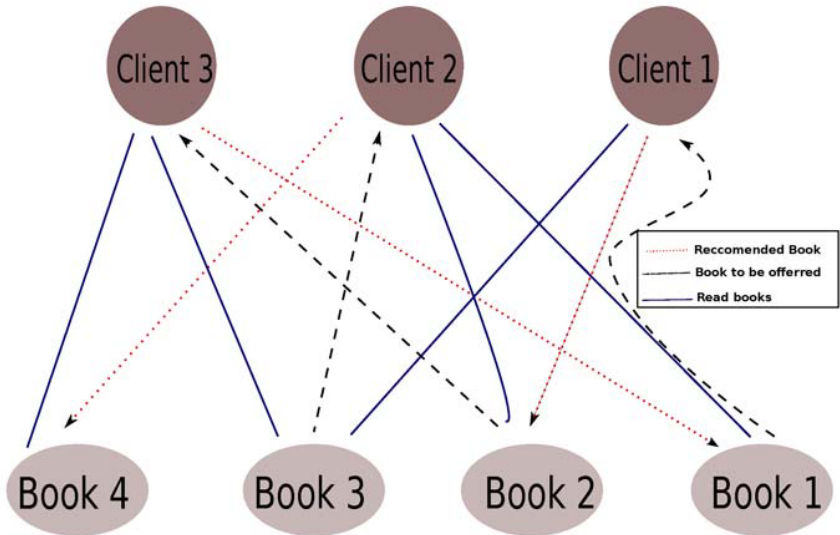


Fig. 1. Quasi-Biclique with books to be offered and recommended

AI mining tools associated with marketing response models are changing the central role of information previous labeled as useless. Datasets generated over a long period of time, or even short time ones with high measuring rate, are being mined and sought for optimization. The TOPDMC is an special case of DM campaigns, where OR is used to optimize the profit from the values generated by those response models.

Given a set of clients $C = \{c_1, \dots, c_m\}$ and a set of offers $B = \{b_1, \dots, b_n\}$, a cost d_{ij} and profit r_{ij} are associated to each book offer $j \in B$ targeted to a specific customer $i \in C$. Those estimated parameters d_{ij} and r_{ij} are usually obtained from the aforementioned marketing response models. For each client

$i \in C$, there is a number of maximum offers M_i , which indicates the maximum number of offers that a client i is willing to receive. For each product offer $j \in B$, there is a strict minimum number of products O_j^{min} that should be offered during the campaign, a minimum expected profit R (also known as hurdle rate), an available budget B_j and, finally, a fixed cost f_j if the product j is chosen to be disseminated in that campaign.

The 0-1 Multiple Knapsack Problem is a special case of the TOPDMC. In this sense, since the Multiple Knapsack Problem belongs to the $\mathcal{NP}$ -Hard class, the TOPDMC also does. Different mathematical formulations were exploited, and compared, in the work of Nobibon et al. [5].

2 Quasi-Bicliques generation and probabilities extractions

The generator consider bipartite graphs that have been clustered before in some disjoint bicliques. This problem could be solved by finding the minimum editing distance in order to find clustered bicliques, that become quasi-bicliques when these extra edges are removed. Although the clustering problem is NP-hard itself, we manage to generate instances in polynomial time. The process is described in the following steps:

- (i) Start with a fully connected biclique graph (clients versus books);
- (ii) Randomly choose clients and books in order to create bicliques;
- (iii) Randomly remove edges from each biclique maintaining the connectivity (respecting at least *maxClientsPerBiclique* and *maxBooksPerBiclique* per biclique), thus forming quasi-bicliques;
- (iv) Calculate the probability for each pair (client, book), after the insertion of an extra edge.

The calculation of the probability is done as follows. Given a new edge (c, b) not present in the quasi-biclique (C, B) containing the client c , insert this edge in the quasi-biclique. The probability of client c to buy another recommended book b' is given by $\delta(b')/|B|$, that is, the more connected the clients and books are, bigger is the value. Finally, remove the extra edge in order to get the original quasi-biclique with the probability values stored in matrix $p_{b,c,b'}$. Given the sum of the probabilities of each book b' multiplied by the profit of the book minus the production cost, you get the estimated revenue for the investment represented by edge (c, b) . Thus, the same index r_{ij} defined for the TOPDMC can be calculated as defined in Eq. (1), where pp is

the profit percentage estimated by the book seller and c and b are, respectively, clients i and j from the previous description.

$$(1) \quad r_{cb} = \sum_{b'=1}^n p_{b,c,b'}(d_{ij}pp - d_{ij})$$

2.1 Book campaigns instance generator

Given the 3D matrix of probabilities $p(b, c, b')$, the cost of offering a book were generated at random, $d_{ij} = bc_j + dc_i$, with $bc_j \in [10, 100]$ and $dc_i \in [5, 25]$, where bc_j represents the cost of manufacturing and paying the copyright of book j while parcel dc_i represents the delivering cost of offering some book to client i . The return to the firm r_{ij} is described in Eq. (1).

The campaign operational constraints were generated following the strategy used by Nobibon et al. [5]. Thus, the minimum-quantity commitment bound O_j (the minimum amount of books that should be manufactured in the marketing campaign) was generated as a random integer selected between $\lfloor \frac{\sum_i M_i}{n} \rfloor$ and $\lfloor 2 \frac{\sum_i M_i}{n} \rfloor$. The budget B_j was chosen between $O_j \frac{\sum_i c_{ij}}{m}$ and $2 \frac{\sum_i c_{ij}}{m}$. However, all fixed cost f_j were chosen to be 1. Campaign hurdle rate was also fixed in 15%.

A set of 75 instances was generated considering campaigns with 100 clients and the following number of available books: 10, 20, 50, 100 and 200. The rate of profit for each book was also objective of our analyzes, considering the following values: $pp = [110\%, 120\%, \dots, 250\%]$. Costs were fixed for instances with the same number of books. Thus, the only changes among them are the revenue for each offer. Parameter *maxClientsPerBiclique* was set to be number of clients divided by 3 and *maxBooksPerBiclique* was set as half of the number of books.

3 TOPDMC optimization using a VNS algorithm

3.1 Solution representation and evaluation

A simple way to represent a solution to the TOPDMC is thought an array $R_{|C| \times |O|}$ of binary variables, in which C indicates the set of available customers able to receive offers and O , representing the set of products quoted for the marketing campaign. If a given cell $s_{i,j} | i \in C, j \in O$ is true, the product j is going to be offered for client i .

A given solution s is evaluated by measuring the total profit of the campaign, which represents the total expected profit minus the total costs. The

latter is a combination of individual clients costs plus fixed costs of each product. Thus, the goal is to maximize the evaluation function $f^{obj}(s)$ (Eq. (2)).

$$(2) \quad f^{obj}(s) = \sum_{i \in C} \sum_{j \in P} (r_{ij} - d_{ij}) s_{i,j} - \sum_{j \in P} f_j y_j$$

3.2 Neighborhood structures

To explore the search space of the TOPDMC, three Neighborhood Structures (NS), proposed by Nobibon et al. [5] and also used by Oliveira et al. [6], are considered in this study:

Swap Clients Intra – $NS^{SC^{Intra}}(s)$: This move swaps two clients, one active and another not offered yet, $l, m \in C$ of a given product $j \in O$, such that $s_{l,j} = s_{m,j}$ and $s_{m,j} = s_{l,j}$.

Swap Clients Inter – $NS^{SC^{Inter}}(s)$: Similar to the previous one, but, in this case, offers from different books $i, j \in O$ are swapped, e.g., $s_{l,i} = s_{m,j}$ and $s_{m,j} = s_{l,i}$.

Swap Products – $NS^{SP}(s)$: exchanges two different columns $i, j \in O$ of a given solution s , such that $y_i = 1$ and $y_j = 0$ or unlike. The now book is only added to the marketing campaign if there is, at least, O_j clients available.

3.3 GGVNS with RDM local search

The optimization of the remaining TOPDMC, after generating the book marketing campaign, is done by a previous designed metaheuristic algorithm [6], combining the GRASP and GVNS [2]. The GRASP construction phase uses a strategy improved from the literature and introduced by Oliveira et al. [6]. This greedy randomized solution generated returns the best solution found among $GRASP_{MaxIter}$ generated solutions.

The GVNS exploits the search space by using the aforementioned NS. The local search was done using three Random Descent Methods (RDM), without guarantee of local optimality. A RDM is created for each neighborhood and a maximum number of random moves m_i^{RDM} are applied, for each NS i . The RDM procedure consists in randomly extracting solutions from a given NS, determining its objective function value, and moving to that solution if improvement is found. When a better solution is found, an auxiliary counter returns to 0, otherwise, it keep going until m^{RDM} random are done without any improvement. Finally, a Variable Neighborhood Descent [4] is created with random order of the NS. The idea of using a VND based on random

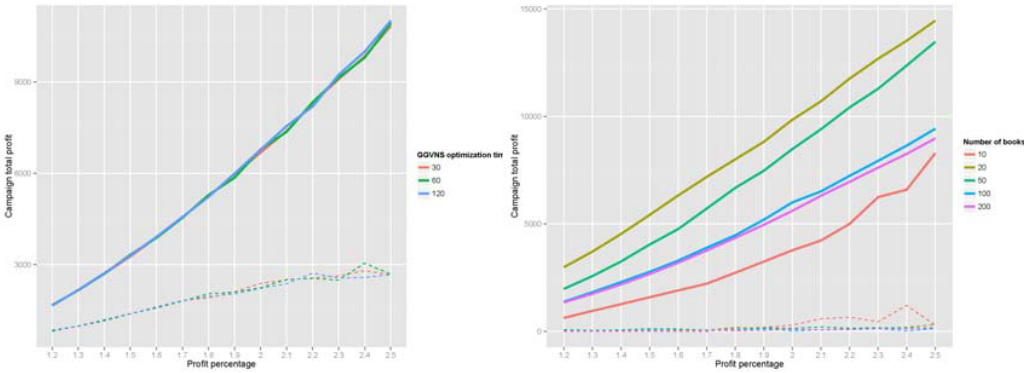
descents reduces the computational cost of the local search, however, consequently reducing its ability of finding local optimum solutions after its search.

4 Computational experiments and discussion

The optimization GGVNS algorithm was implemented in C++ in the framework OptFrame 2.2³. The tests were carried out on a Notebook Intel i7-3537U 2.00GHz, DDR3 1.6 GHZ with 8GB of RAM, with operational system Ubuntu 14.04 and compiled by g++ 4.8.4.

4.1 GGVNS results

The GGVNS algorithm was applied for solving each of the 75 instances, considering executions of 30, 60 and 120 seconds. Since campaigns expected profit (HR) were fixed in 15%, and modeled as a hard constraint, some instances with $pp = 110\%$ were not feasible. The feasibility is possible in some configurations due to the fact that by offering a given book b may result in a client buying several other recommended books b' . However, for simplicity, we removed these five instances from the following analyzes.



(a) Book profit percentages and GGVNS optimization time (b) Book profit percentages and number of books

Fig. 2. Interactions plots and model classification accuracy

Figure 2 shows an interaction plot regarding book profit percentage, expected campaign profit and GGVNS optimization time. The dashed line indicates the standard deviation while the thicker line shows the average total

³ Available at <http://sourceforge.net/projects/optframe/>

campaign profits. As can be verified, due to small size of the generated instance, the GGVNS may have reached solutions close to optimality and reported similar values for all analyzed computational times. When books profit rate are higher (more than 170%), campaign total profit seems to increase quicker. This fact can be associated with the higher number of possible profitable clients to be chosen in the marketing campaign.

5 Final considerations and possible extensions

In this brief manuscript, we introduced a novel direct marketing campaign that deals with free book sampling. In particular, the proposed approach could be also applied in the cosmetic sector, where free samples are given to costumers. A new set of instances, comprising large sets of clients and available books should be carefully generated and analyzed.

Acknowledgement

The authors thank the Brazilian agencies CNPq (grants 306694/2013-1, 552289/2011-6 and 202380/2012-2) and FAPEMIG (grants PPM CEX 772/15 and APQ-04611-10) for supporting the development of this work. Helena Ramalhinho was also partially supported by the Spanish Ministry of Economy and Competitiveness (TRA2013-48180-C3-P, TRA2015-71883-REDT).

References

- [1] Amor, I. B. and F. Guilbert, *Influences on free samples usage within the luxury cosmetic market*, Direct Marketing International Journal **3** (2009), pp. 67–82.
- [2] Hansen, P., N. Mladenovic and J. A. M. Pérez, *Variable neighborhood search*, European Journal of Operational Research **191** (2008), pp. 593–595.
- [3] Lammers, H. B., *The effect of free samples on immediate consumer purchase*, Journal of Consumer Marketing **8** (1991), pp. 31–37.
- [4] Mladenovic, N. and P. Hansen, *A variable neighborhood search*, Computers and Operations Research **24** (1997), pp. 1097–1100.
- [5] Nobibon, F. T., R. Leus and F. C. Spieksma, *Optimization models for targeted offers in direct marketing: Exact and heuristic algorithms*, European Journal of Operational Research **210** (2011), pp. 670–683.
- [6] Oliveira, T. A., V. N. Coelho, M. J. F. Souza, D. L. T. Boava, F. Boava, I. M. Coelho and B. N. Coelho, *A hybrid variable neighborhood search algorithm for targeted offers in direct marketing*, Electronic Notes in Discrete Mathematics **47** (2015), pp. 205–212.