OPTIMIZED DELIVERY OF ON-LINE ADVERTISEMENTS
A Linear Programming Approach to the Delivery of On-line Advertisements

Fabrizio Caruso[1] and Giovanni Giuffrida[2]
[1]*Neodata, Catania, Italy*
[2]*Department of Social Sciences, University of Catania, Catania, Italy*

Keywords: Web advertisement, Linear programming, Data mining, Machine learning.

Abstract: We find an optimal strategy for displaying advertisements in given locations at given times under some realistic dynamic constraints. Our goal is to maximize the total profit produced by the impressions, which depends on profit-generating events such as the impressions themselves and the ensuing clicks. We must take into account the possibility that the constraints could change over time in a way that cannot always be foreseen.

1 INTRODUCTION

We want to find the optimal strategy for displaying advertisements in order to achieve different goals (maximum total profit, maximum visibility of the campaign, etc.) under some realistic constraints. We need to find the optimal number of "impressions" (display of advertisements at a given time and at a given "location") under some realistic dynamic constraints that both limit the possibility of certain creatives and limit the number of impressions in certain locations and/or moments in time. A location is a place where an advertisement can be displayed. This model can be generalized to target users by their categories, as well. Similar optimization problems have been considered in the literature (Vee et al., 2010; Alaei et al., 2009; Abrams et al., 2007; Langheinrich et al., 1999; Abe and Nakamura, 1999; Nakamura, 2002). Other approaches have also been considered by the authors of this article (Caruso et al., 2011), where a Bayesian model is used.

Our approach improves over the previous ones in different respects:

- we consider a more realistic model (realistic constraints of different nature);
- we formally investigate the problem of consistency of the constraints;
- we consider an optimization of the problem by approximating it to a problem with fewer unknowns;
- we can apply machine learning techniques for guessing the traffic on locations.

We are given certain "creatives" (advertisements, e.g. banners, videos, etc.), "campaigns" (sets of related creatives), certain "locations" and a period of time (set of "time frames"). At a given moment in time we have an expected profit for each creative of a given campaign in a given time and location.

The profit of the web-page's owner depends on the profit-generating events that have been agreed upon by the advertiser and the web-page's owner. These events can be the impression itself, a click on the advertisement or a registration of any sort (e.g. registration into the advertised site, purchase of the advertised item, etc.), or any combinations of these events. We denote the expected profit of a single "impression" as the "impression profit" (the expected profit of a single impression obtained by all the profit-generating events such as the impression itself, the ensuing click and registrations of all types). In such a way we can avoid keeping track of click-through rates and different registration rates. This choice is a compromise between performance and generality, since it makes our model less precise and slightly less general: we are not considering campaigns with separate budgets for different events; we cannot estimate the expected profit of an impression as precisely as when different rates for different events are considered.

The number of impressions on a given location at a given time is limited by the traffic ("supply") of the corresponding webpage. It also depends on time in a way that can be only partially predicted. Moreover the maximum profit for a given campaign ("demand") may be limited by a predefined budget.

Our goal is to maximize our expected revenue

Caruso F. and Giuffrida G..
OPTIMIZED DELIVERY OF ON-LINE ADVERTISEMENTS - A Linear Programming Approach to the Delivery of On-line Advertisements.
DOI: 10.5220/0003805504000405
In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence* (ICAART-2012), pages 400-405
ISBN: 978-989-8425-95-9

which is given by the expected total price paid.

Therefore we wish to maximize a weighted sum of all expected profits obtained in all locations in the period of time under consideration.

Taking into account only supply and demand constraints makes our model a special instance of a "transportation problem" for which very efficient solutions exist (see (Dantzig, 1963)).

The complexity of the model brings up the additional problem of deciding between simplifying the model and considering smaller problems.

In order to apply our optimization we need to make a projection of the future supply and a projection of the impression profits onto our period. Impressions are only possible on the locations and times allowed by the scheduling. The projection of the impression profits should also try to "guess" how the profit of an impression changes in time. The projection algorithms should take into account different periodicities (e.g., daily, weekly). The projections can be improved by applying machine learning techniques to compute the weights of the periodicities.

Moreover we cannot assume the immutability of the constraints of the problem in the period of time under consideration. For this reason we have to continuously readjust to new conditions.

2 NOTATION

We denote by C_i the i-th campaign (set of creatives) and by $B_{i,j}$ its j-th creative, by L_l the l-th location, by T_k the k-th time frame.

We denote the "impression count" by $x_{i,j,k,l}$, i.e., the number of impressions of $B_{i,j}$ at time frame T_k and at location L_l. For example $x_{2,3,1,5} = 10$ means that banner $B_{2,3}$ is displayed ten times at time T_1 at location L_5.

We denote by $p_{i,j,k,l}$ the "impression profit", the profit of $B_{i,j}$ at location L_l and a time T_k.

2.1 Configurations

We can consider our problem as the problem of finding the optimal impression counts for the entries in a tridimensional matrix, i.e. the points in a discrete finite space given by a grid defined by couples (campaign, creative), time and location.

We refer to a single point (i, j, k, l) in this tridimensional discrete finite space as an "impression-event" (or simply an "impression" when this is clear from the context). We call any choice for the values of $x_{x,j,k,l}$ of all the impression-events as a "configuration". An impression is in fact characterized by a couple (campaign, creative), a location and a time. Our goal is to choose the optimal delivery of each possible impression, i.e. an optimal configuration. We will simply refer to the the number of impressions of an impression-event as the "impression count".

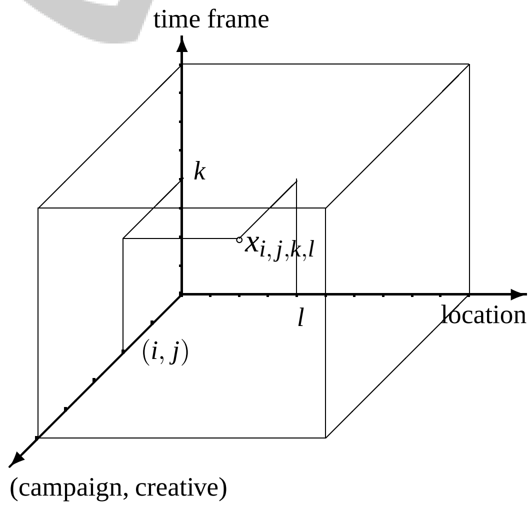
We refer to the points that are allowed by the schedule as to "admissible points". Each admissible point describes a dimension of our optimization problem. The worst case is produced when all points inside the cube of size given by the number of couples (campaign, creative), the number of time frames, the number of locations, are admissible. Hence in the worst case the number of dimensions is the product of the number of couples (campaign,creative), the number of locations and the number of time frames considered.

Remark 1. *In practice we need to translate the number of impressions $x_{i,j,k,l}$ in terms of probability of delivery. We transform a configuration into a map*

$$(i, j, k, l) \mapsto Prob_{k,l}(i, j)$$

where $Prob_{k,l}(i, j)$ is the computed probability of delivery of $B_{i,j}$ at location L_l and time T_k, in that we take the ratio between $x_{i,j,k,l}$ and the expected supply $\mathcal{S}_{l,k}$ at L_l and time T_k.

Pictorially, we could see a single configuration $C = (x_{i,j,k,l})_{i,j,k,l}$ as a tridimensional matrix:



3 REALISTIC MODEL

We want to consider a realistic model in which several constraints of different nature are taken into account.

3.1 The Constraints

We distinguish between the primary (physical) constraints of the problem, the secondary ones (commercial and optional) and the learning constraints (required by the learning phase if it is included in the mathematical model).

3.2 Primary Constraints

The primary constraints are given by the schedule of the campaigns, by the limited supply of impressions and by a (possibly) limited demand (budget):

1. The scheduling of the campaigns limits the admissible points: certain creatives $B_{i,j}$ are only possible at certain times and locations. Typically a campaign begins and ends at certain times and its creatives are limited to certain locations, hours of the day, days of the week, etc.
2. Any location at a given time receives a limited supply of impressions, which solely depends on the traffic of its page;
3. For any given campaign a given total profit may not be exceeded ("demand") because only a finite campaign budget can be available.

Remark 2. *Campaigns can have an unbounded budget, e.g. one that only pays for an actual purchase.*

3.3 Secondary Constraints

The secondary constraints may be of a commercial nature. They could be enforced in real time while monitoring the delivery, although having them as constraints improves the accuracy of the model. They are necessary to increase the visibility of a certain campaign/creative:

1. Any given creative/campaign should not last less than a given period, e.g. the period in which the campaign is scheduled. We enforce this by setting a minimum for the number of impressions for each possible time frame.
2. We would like to avoid having only one creative at a given location and time frame when more than one choice is available.

3.4 Learning Constraints

We can embed some learning constraints into the mathematical model. One way to achieve this can be a constraint of the form: for each new couple $(creative, location)$ we must have a minimum number of impressions in all (or some initial) possible time frames. This is not strictly necessary because the same goal can be achieved by using portion of the traffic for learning. Having these constraints inside the model produces a more accurate model.

4 LINEAR PROGRAMMING

Under the mild hypothesis that the impression profit $p_{i,j,k,l}$ is constant with respect to its count $x_{i,j,k,l}$ we can assume that our constraints are linear. This assumption is not true in general because there is no linear dependence between the total profit generated by an impression-event and an impression count, i.e. displaying the same advertisement x times on the same location, possibly more than once to the same user, does not necessarily produce x times the profit produced by one single display.

Since we are ultimately interested in the probability of delivery and since integer linear programming is computationally unfeasible (NP-hard), a possible approach to this problem could be real linear programming: we approximate our discrete problem with a continuous one and we do not mind considering a real number of impressions.

4.1 Formalized Constraints

The points that do not contradict the first primary constraint will be the unknowns of our model.

4.1.1 Primary Constraints

We do not include the first primary constraints for the reasons given above and assume that in our expressions all indices run over points that do not contradict the first primary constraints.

Supply and demand are formalized as follows: Second primary constraint:

$$\forall_{l,k} \sum_{i,j} x_{i,j,k,l} \leq \mathcal{S}_{l,k}; \text{ (supply)} \tag{1}$$

where $\mathcal{S}_{l,k}$ is the supply at location L_l and at time T_k.

Third primary constraint:

$$\forall_i \sum_{j,l,k} p_{i,j,k,l} x_{i,j,k,l} \leq \mathcal{D}_i; \text{ (demand)} \tag{2}$$

where $\mathcal{D}_i$ is the budget of the i-th campaign.

Remark 3. *If only the primary constraints are taken into account, we have a "Hitchcock's style transportation problem" (Hitchcock, 1941). For such problems very efficient algorithms are known such as the "stepping stone algorithm" (Dantzig, 1963).*

4.1.2 Secondary Constraints

The secondary constraints are formalized as follows: First secondary constraint:

$$\forall_{i,k} \sum_{j,l} x_{i,j,k,l} \geq \mu_{i,k}; \text{ (duration)} \tag{3}$$

where $\mu_{i,k}$ is the desired minimum delivery of impressions of the i-th campaign at time T_k.

Second secondary constraint:

$$\forall_{l,k\in\mathcal{D}}\forall_{i,j}\ x_{i,j,k,l} \leq P_{l,k}\cdot \mathcal{S}_{l,k};\ \text{(no overflow)} \qquad (4)$$

where $P_{l,k} \in [0,1]$ (usually close to 1) defines how much a single creative can occupy a location at a given time frame and where $\mathcal{D}$ is the set of indices corresponding to locations and time frames where at least 2 different creatives are possible.

Remark 4. *The second secondary constraints (4) should only be limited to those cases in which at a given location and time more than one pair of campaign and creative is possible because otherwise the constraint would prevent the location from being filled with impressions even when this could be possible.*

4.1.3 Learning Constraints

If the learning phase is included in the model some constraints should force a minimum delivery for the new creatives and new locations:

$$\forall_{\text{new } j,l}\forall_{i,k}\, x_{i,j,k,l} \geq \lambda_{i,j,k,l}. \qquad (5)$$

We are also implicitly assuming that the unknowns are non-negative, i.e.

$$\forall_{i,j,k,l}\ x_{i,j,k,l} \geq 0. \qquad (6)$$

4.2 The Objective Function

We want to maximize our expected revenue, which is given by the sum of all expected profits received in a given configuration C:

$$F(C) = \sum_{i,j,k,l} p_{i,j,k,l}\, x_{i,j,k,l}. \qquad (7)$$

where $p_{i,j,k,l}\, x_{i,j,k,l}$ is the expected profit generated by $B_{i,j}$ at location L_l and at time T_k.

5 EXISTENCE OF A SOLUTION

We see that there is no guarantee of consistency once the secondary and learning constraints are introduced, even if we exclude the first primary constraints. We need to solve the system of inequalities in order to know if there is a solution. Nevertheless we can use some heuristic method to avoid some inconsistent problems. In particular we can find some conditions on $\mu_{i,k}$ in (3) and $\lambda_{i,j,k,l}$ in (5) under which the problem cannot have a solution.

Given a set of t-uples T, we introduce the following notation for subsets of $t-1$-uples:

$$\begin{aligned} T[i\to\alpha] &= \{(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_t) \\ &\quad |(x_1,\ldots,x_{i-1},\alpha,x_{i+1},\ldots,x_t)\in T\}; \\ T[i\to *] &= \{(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_t) \\ &\quad |\exists v|(x_1,\ldots,x_{i-1},v,x_{i+1},\ldots,x_t)\in T\}. \end{aligned}$$

i.e. we are considering respectively

- $(t-1)$-uples obtained from t-uples in T, where i-th component is α and it has been removed;
- $(t-1)$-uples obtained from t-uples in T where the i-th component has been removed independently of its value.

In the same way, if more components are removed in parallel, we introduce the notation: $T[i_1 \to \alpha_1,\ldots,i_n \to \alpha_n]$ for $(t-n)$-uples, where $\alpha_j \in \mathbb{N}\cup\{*\}$ and $i_j \in \mathbb{N}$ for $j \in \{1,\ldots,n\}$. Here we will consider the set T of all indices (i,j,k,l) that are allowed by the first primary constraints.

Examples.

$$\begin{aligned} &T[1\to i,2\to *,4\to *] = \{k \mid \exists j,l \mid (i,j,k,l)\in T\} \\ &T[1\to i] = \{(j,k,l) \mid (i,j,k,l)\in T\} \\ &T[3\to k,4\to l] = \{(i,j) \mid (i,j,k,l)\in T\} \\ &T[2\to *,3\to *,4\to *] = \{i \mid \exists j,k,l \mid (i,j,k,l)\in T\} \\ &T[1\to *,2\to *] = \{(k,l) \mid \exists i,j \mid (i,j,k,l)\in T\} \end{aligned}$$

Fact 1. *Let us assume $T[1\to i,2\to *,4\to *]\neq\emptyset$ for all $i\in T[2\to *,3\to *,4\to *]$ (i.e. for all possible campaigns C_i).*

If we choose

$$\mu_{i,k} > \frac{\mathcal{D}_i}{|T[1\to i,2\to *,4\to *]|m}$$

where $m = \min p_{i,j,k,l}$, then the semi-algebraic set defined by the inequalities (2), (3) is empty.

Proof. By the first secondary constraint (3) we have

$$\sum_{j,l} x_{i,j,k,l} > \frac{\mathcal{D}_i}{|T[1\to i,2\to *,4\to *]|m}.$$

Therefore for any campaign C_i we have

$$\begin{aligned} &\sum_{j,k,l} p_{i,j,k,l}x_{i,j,k,l} \geq \\ &\geq m\sum_{j,k,l} x_{i,j,k,l} \geq m \sum_{k\in T[1\to i,2\to *,4\to *]}\ \sum_{j,l} x_{i,j,k,l} > \\ &> \sum_{k\in T[1\to i,2\to *,4\to *]} \frac{\mathcal{D}_i}{|T[1\to i,2\to *,4\to *]|} = \mathcal{D}_i. \end{aligned}$$

□

Fact 2. *Let us assume* $T[1 \to i], T[3 \to k, 4 \to l] \neq \emptyset$ *for all* $i \in T[2 \to *, 3 \to *, 4 \to *]$ *(i.e. all possible campaigns* C_i*) and* $(k,l) \in T[1 \to *, 2 \to *]$.

If all banners and locations under consideration are new and we choose

$$\lambda_{i,j,k,l} > \min\left\{\frac{\mathcal{D}_i}{|T[1 \to i]|m}, \frac{\mathcal{S}_{k,l}}{|T[3 \to k, 4 \to l]|}\right\}$$

where $m = \min p_{i,j,k,l}$, *then the semi-algebraic set defined by the inequalities (1), (2), (5) is empty.*

Proof. If $\lambda_{i,j,k,l} > \frac{\mathcal{D}_i}{|T[1\to i]|m}$ then

By (5) we have

$$x_{i,j,k,l} > \frac{\mathcal{D}_i}{|T[1 \to i]|m}$$

from which it follows that for any campaign C_i:

$$\sum_{j,k,l} p_{i,j,k,l} x_{i,j,k,l} \geq m \sum_{j,k,l} x_{i,j,k,l} >$$

$$> m \sum_{j,k,l} \frac{\mathcal{D}_i}{|T[1 \to i]|m} = \mathcal{D}_i.$$

If $\lambda_{i,j,k,l} > \frac{\mathcal{S}_{k,l}}{|T[3\to k,4\to l]|}$ then by (5) we have

$$x_{i,j,k,l} > \frac{\mathcal{S}_{k,l}}{|T[3 \to k, 4 \to l]|}$$

from which it follows that for any couple $(k,l) \in T[1 \to *, 2 \to *]$ we have

$$\sum_{i,j} x_{i,j,k,l} > \sum_{i,j} \frac{\mathcal{S}_{k,l}}{|T[3 \to k, 4 \to l]|} = \mathcal{S}_{k,l}.$$

□

6 FORECASTING DATA

In order to apply our optimization algorithms we need to have at least a projection of the supply and a projection of the expected profit of all impressions allowed by the first primary constraint. The supply and impression profits can be estimated by taking a proper weighted average from the historical data. The projection should take into account different factors: episodic factors and possibly different periodicities.

6.1 Projecting the Profit

The profit of an impression-event may depend on the periodicity of its campaign and of its location. Since the "impression profit" changes slowly in time, it can be predicted better than the supply. If we want to determine the expected profit for an impression-event we can take some average profit from historical data on "similar" events. Our strategy is to use the most accurate and recent available information.

6.2 Projecting the Traffic

A model for the projection of the supply should take into consideration the periodicity of the location, i.e. some sites are more often visited in particular periods of the year, day, hours, etc. More periodicities may concur, e.g. a site may be visited more often in a specific day of the week and at a specific hour of the day. Regression analysis through machine-learning techniques such as support vector machines can be a viable approach for the problem of properly choosing the weights of the average of the different "features".

7 FURTHER IMPROVEMENTS

This approach can be improved in its accuracy by targeting the users, and in its speed by reducing the dimensions and constraints in the model.

7.1 Targeting Users

The approach we have presented optimizes the delivery of advertisements in both space (locations) and time (time frames). The very same algorithms and code can be used to take users' profiles into account by encapsulating the profile information into the information on the location by storing a pair $(location, profile)$ into a single "extended location".

7.2 Simplifying Things

The large number of unknowns and constraints in this general approach can pose a serious problem to its computable feasibility. We can reduce the dimensions and constraints by clustering similar attributes, (Abe and Nakamura, 1999) or by simplifying our model:

- We restrict our problem to periods of time in which the time constraints do not change. This greatly reduces the number of unknowns but could produce suboptimal solutions.
- We avoid secondary and learning constraints and enforce them during the delivery.
- We use a time horizon, beyond which all the time frames are considered as a single time.

8 RESULTS ON REAL DATA

We have implemented an ad-server optimizer in `Java`. For solving the linear programming model we use `glpk`[1]. Optionally we use the freely available sup-

[1] http://www.gnu.org/s/glpk/

port vector machines library[2] to project future supply.

Our implementation requires as input: historical data necessary for projecting the impression profits and the future supply, campaign data (budgets), scheduling data.

Our prototype has been used on real data available at Neodata and has been compared against the results of the currently used optimizer, which works as follows: if a campaign is achieving its target at the current rate, nothing is done, otherwise, it is stopped in its less profit-generating locations.

We used logs and schedules of two clients of Neodata, which, we call *A* and *B*. We remark that the traffic managed by Neodata, neither accounts for the total traffic nor is it a constant percentage of the traffic generated by the sites under consideration. *A* was optimized equally well by the current optimizer and our prototype; whereas *B* was optimized better by a large margin (20% - 50%) by our code. In the following table we show the result of one of our experiments on the data of April 30th 2010 for company *B*:

hour	real profit	opt. profit	gain	% gain
8:00 a.m.	2.50	4.41	1.91	76%
9:00 a.m.	3.96	8.07	4.11	104%
10:00 a.m.	6.69	12.97	6.28	94%
11:00 a.m.	14.17	23.32	9.15	65%
12:00 a.m.	14.98	24.66	9.68	65%
1:00 p.m.	15.00	14.01	-0.99	-7%
2:00 p.m.	19.43	31.81	12.38	64%
3:00 p.m.	26.07	41.14	15.07	58%
4:00 p.m.	23.38	24.37	0.99	4%
5:00 p.m.	13.98	14.40	0.42	3%
6:00 p.m.	12.64	28.74	16.10	127%
7:00 p.m.	15.90	28.38	12.48	78%
8:00 p.m.	10.55	10.89	0.34	3%
total	179.25	267.17	87.92	49%

Possible reasons why data on *A* are not optimized equally well may be: there is no room for further improvement; the data on the supply cannot be used for the projection because it does no correspond to a constant percentage of the real traffic.

The data was used as follows: the initial portion of the data (e.g. the first 20 days) were used for training the system, i.e. projecting the supply (traffic) and the profits. The remaining part of the month was used as a schedule and was optimized.

9 CONCLUSIONS

We have shown how a linear programming approach can be used to optimize the delivery of on-line banners. Our approach takes many different constraints into account (schedule, supply, demand, visibility, learning). We prove that under some conditions, the corresponding system of inequalities is consistent.

This approach can be used to target users by simply extending the concept of location. We have also tackled the problem of dimensionality (by a time horizon, by simplifying the constraints, etc.).

Our prototype has been tested on real data. We have shown that it optimizes the delivery of on-line advertisements better than a greedy algorithm.

There are still some open issues: how to project the traffic when the conditions of the problem change quickly and the data does not correspond to a constant percentage of the traffic.

[2] http://www.csie.ntu.edu.tw/˜cjlin/libsvm

REFERENCES

Abe, N. and Nakamura, A. (1999). Learning to Optimally Schedule Internet Banner Advertisements. In *ICML '99: Proceedings of the Sixteenth International Conference on Machine Learning*, pages 12–21, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Abrams, Z., Mendelevitch, O., and Tomlin, J. (2007). Optimal delivery of sponsored search advertisements subject to budget constraints. In *Proc. of the 8th ACM conf. on Electronic commerce*, EC '07, pages 272–278, New York, USA. ACM.

Alaei, S., Arcaute, E., Khuller, S., Ma, W., Malekian, A., and Tomlin, J. (2009). Online allocation of display advertisements subject to advanced sales contracts. In *Proc. of the 3rd International Workshop on Data Mining and Audience Intelligence for Advertising*, ADKDD '09, pages 69–77, New York, USA. ACM.

Caruso, F., Giuffrida, G., and Zarba, C. (2011). Real-time Behavioral Targeting of Banner Advertising. In *Book of Abstracts of the CLADAG 2011*.

Dantzig, G. (1963). *Linear Programming and Extensions*. Princeton University.

Hitchcock, F. (1941). The distribution of a product from several sources to numerous localities. *Journal of Mathematics and Physics*, 20:224–230.

Langheinrich, M., Nakamura, A., Abe, N., Kamba, T., and Koseki, Y. (1999). Unintrusive customization techniques for web advertising. *Computer Networks*, 31(11-16):1259–1272.

Nakamura, A. (2002). Improvements in practical aspects of optimally scheduling web advertising. In *WWW '02: Proc. of the 11th international conf. on World Wide Web*, pages 536–541, New York, USA. ACM.

Vee, E., Vassilvitskii, S., and Shanmugasundaram, J. (2010). Optimal online assignment with forecasts. In *Proc. of the 11th ACM conf. on Electronic commerce*, EC '10, pages 109–118, New York, USA. ACM.