MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
SUMY STATE UNIVERSITY
UKRAINIAN FEDERATION OF INFORMATICS

PROCEEDINGS

OF THE V INTERNATIONAL SCIENTIFIC CONFERENCE

ADVANCED INFORMATION SYSTEMS AND TECHNOLOGIES

AIST-2017

(Sumy, May 17–19, 2017)

SUMY
SUMY STATE UNIVERSITY
2017

Optimization of Websites Advertising Budget Distribution

Pavel Mishchenko, Serhey Shapovalov
Sumy State University, Ukraine, shap@id.sumdu.edu.ua

***Abstract* – This paper is devoted to problem of the most optimal allocation of website`s advertising budget among teaser networks. We consider the possibility of improving the efficiency of distribution by means of an automated system. The main steps of genetic algorithm and integral estimation adaptation to the multiple-criteria problem solution are shown.**

***Keywords* – Teaser Network, Genetic Algorithms, Integral Estimation, Optimal Distribution.**

I. Introduction

A significant share of commercial website budget (e.g. online stores etc.) refers to advertising and customer acquisition (Table I).

TABLE I. BUDGET DISTRIBUTION BY ADVERTISING DIRECTIONS

Direction	**Annual results, million UAH**	
	2015	***2016***
Search (paid issue in search engines)	880	990
Banner advertising	680	710
Mobile advertising	150	200
Digital video	325	455
Total market	**2 035**	**2 355**
Another digital	200	230
Total	**2 235**	**2 585**

Vast expansion of services provision to Internet users forces to look for new ways of overcoming their irritations associated with obsessive advertising or various referral links to follow. Teasers – the next generation of banner and contextual advertising – is one of the possible solutions. A teaser is an advertising message, which contains an attention-grabbing picture and a short description, like an intriguing title [3]. To place teaser advertisement on different platforms, websites apply to services of teaser networks (TN) – intermediaries, which are responsible for advertisements allocation and perform a quality control of targeted platforms.

II. Problem Statement and Description of the Mathematical Model

The Internet advertising market is very widespread (Table II). Thus, a choice of TN is a key mission of promotion specialists. Commonly, they place advertising on several teaser networks. Then they analyze the results of their decision for a certain period, and try to find the optimal option of the budget distribution [1]. Such method is extremely time-consuming, it involves a large number of routine and monotonous tasks, that can finally lead to analytical errors and influence of personal preferences in context of human factor. Altogether, this approach negatively affects the number of new website visitors and the total cost of advertising. Therefore, to increase the number of visitors in terms of limited budget we need another tool for analyzing and allocating the budget between the TNs.

TABLE II. TEASER NETWORKS RATING

Position	**Teaser network**	**Year of foundation**	**Traffic, mln clicks / day**
1	AdHub	2010	90
2	Visit Web	2009	60
3	Trafmag	2012	60
4	Recreativ	2009	50
5	TeaserNet	2008	7

The purpose of this study is to optimize the distribution of the limited advertising budget of commercial website B_{sum} by means of automated system. The effectiveness of each TN will be defined as:

$$0 \le E = \frac{k_c}{k_v} \le 1,$$

where k_c – amount of unique visitors, provided by a teaser network, k_v – amount of paid displays of website advertisement. Accordingly, the number of users attracted per day from the particular TN can be calculated by the following formula:

$$k_{di} = \frac{b_{di}}{p_i} * E_i,$$

where d – is the sequence number of the day of research, i – is the sequence number of TN, b_{di} – is a share of daily budget, allocated to certain TN, p_i – is a price of 1 displaying of advertisement inside the TN.

Hence, a model of the optimal budget distribution is formulated:

$$K_{avg} = \frac{\sum_{d=1}^{d_{\max}} \sum_{i=1}^{n} k_{di}}{d_{\max}},$$

where d – is the sequence number of the day of research, i – is the sequence number of TN, d_{max} – is a research duration.

III. Solution of the Problem

In terms of optimizing the budget distribution, it is necessary to take into account the cost of advertisement allocation in various TN and its efficiency. This problem can be considered as a multiple-criteria search problem. To solve it, genetic algorithms can be adapted – a modern and fast tool for finding solutions to multiple-criteria problems.

The first step in applying the genetic algorithm is the choice of chromosomes. It is necessary to distribute the budget among teaser networks t_1, t_2, ..., t_n, therefore the values of chromosome genes will be b_{dj} – starting from B_{sum}/d_{max} to t_i. An additional constraint is $b_{dj} > 0$.

Due to the fact that genes will not be formed classically (usually they are represented by sequence of 0 and 1), it is necessary to change mutation and crossover operations. In order to save the distribution balance, mutation will be performed 2k times (even amount) for each chromosome. The crossover process will be converted into the average value calculation between two corresponding genes of the parent chromosomes

$$(b_{c1}, b_{c2}, ..., b_{cm}) = \left(\frac{b_{b_{11}} + b_{b_{21}}}{2}, \frac{b_{b_{12}} + b_{b_{22}}}{2}, ..., \frac{b_{b_{1m}} + b_{b_{2m}}}{2} \right),$$

where $b_{c1}, b_{c2}, \ldots, b_{cm}$ – are genes of the daughter chromosome;
$b_{b11}, b_{b12}, \ldots, b_{b1m}$ and $b_{b21}, b_{b22}, \ldots, b_{b2m}$ – are genes of the parent chromosomes.

The next significant step is to assess the chromosomes fitness. It is based on calculating the fitness function of each chromosome in a given population. The higher the value of this function - the higher the quality of the chromosome [4].

One of the task`s criteria is the efficiency of the teaser network. The number of times the ad is shown k_v is determined by the share of the budget allocated to the TN, while the number of visitors k_c is estimated based on the website statistics. The effectiveness of the TN is unstable and its dependence on the share of the allocated budget is nonlinear, so the statistics of visits is measured for the last 7 days (Figure 1). Due to the condition $b_i > 0$, we have the following two-dimensional function f(x,t) for each of the teaser networks used.

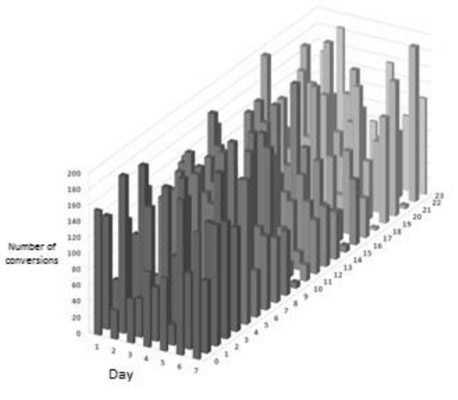

Figure 1. Statistics of conversions from the teaser network

However, the fitness assessment based on a two-dimensional random function in practice is quite complex and time-consuming. It is necessary to use the transition from a two-dimensional random function to a one-dimensional function, by definition of the integral evaluation. Such evaluation Ix(t) for the distribution density f(x,t) of a random variable x was developed in [2]:

$$I(t) = \int_{|x|>|\delta_x|}^{f(x,t)>\varepsilon_x} \frac{1}{|x|} \ln \frac{f(x,t)}{\varepsilon_x} dx,$$

where f(x,t) – the distribution density of a random variable x at time t;

ε_x – a constant parameter having the same dimension as f(x, t). In fact, ε_x determines an unobserved part of generally infinite graph of the distribution density f(x, t);

$|\delta_x|$ – a modulus of the threshold value for x, which was introduced to avoid division by zero. The values ε_x and $|\delta_x|$ are selected by the researcher depending on the problem to solve.

The use of the evaluation makes it possible to quantify the changes in the graph of random variable distribution density or the graph obtained from the histogram observations. Using the integral evaluation, we set the one-dimensional function in accordance with each teaser network. The result is shown at Figure 2.

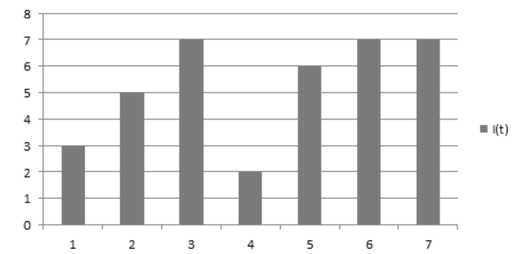

Figure 2. Integral estimation of transitions from a teaser network

As a consequence, the integral evaluation is used as one of the fitness function criteria, since it displays all information about the effectiveness of the TN, but has a simpler way of processing the results. The final fitness function has the following form:

$$\sum_{i=1}^{n} b_i * \sum_{j=1}^{7} I_{ij}(t) \rightarrow \max$$

Chromosomes with the greatest fitness function value will be the most adapted and will participate in breeding. The reason for the break would be the absence of improvement possibility in the solution of the problem. So the best solution is the chromosome with the highest fitness function value. This chromosome is the desired solution of the problem, that is, the vector of characteristics of B_{sum} optimal distribution among the teaser networks.

IV. Results

The study was conducted on the 10 most popular TNs and was limited to 1 calendar month (30 days). Every day the equal part of the budget was distributed among all TNs. With the aim of comparison, the human-expert carried out manual distribution on the similar project with the same budget. The results of manual and automated distribution are available on Figure 3 and Figure 4, respectively. They represent the share in the total budget B_{sum} (in %), allocated to each of the TNs on average for the research period.

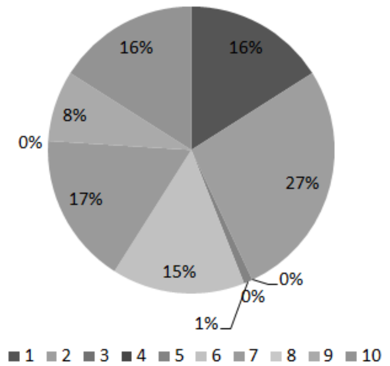

Figure 3. Manual budget allocation

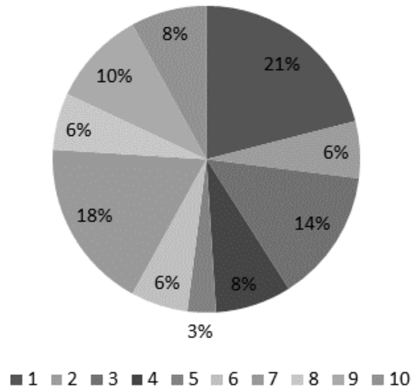

Figure 4. Automated budget allocation

The efficiency of distributions, expressed in the number of unique website visitors, is demonstrated at Figure 5. Horizontal axis represents the sequence numbers of research days (ascendingly). The vertical axis records the statistics of website visits by unique users on the relevant day.

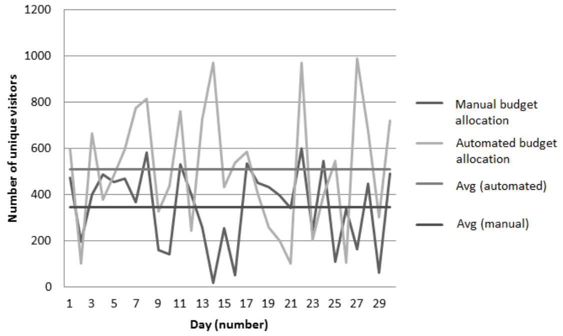

Figure 5. Statistics of site visits by unique users

The results of the automated distribution confirmed the assumption of its advantage in speed of the algorithm's reaction over the expert`s as a whole. Nevertheless, on the chart you can observe 3 days, when the results of manual distribution significantly (by 25-45%) exceeded the automated one. This is because of a problem of incomplete algorithms knowledge in comparison with the expert.

Conclusions

Optimization of budget distribution with the help of a genetic algorithm allowing to maximize the number of transitions to the advertised website in conditions of limited budget. The integral evaluation allowed simplifying and formalizing of the effectiveness estimation for each teaser network used.

The use of automated distribution increased the number of unique visitors in average by 45%, that means, it was 45% more efficient than manual. It is also possible to use automated distribution as a supplementary tool to human decisions. This approach will combine the advantages of an expert and automated methods: implement the expert's knowledge of the outside world, reduce the negative impact of the human factor and increase the speed up the reaction to changes in TN`s effectiveness.

References

[1] Avinash Koshik, *Web 2.0 analytics in practice. Subtleties and best practices.* Dialectics Williams, 2014, p. 218. (in Russian)

[2] V.V. Avramenko, A.P. Karpenko, “Integral estimation of the distribution density of a random variable” in *Modern engineering technologies*, vol. 1. Kiev: ISMO, 1997, p. 263. (in Ukrainian)

[3] Rose Robert., *Content management. A practical guide to creating a loyal audience for your business.* Mann, 2014, p. 57. (in Russian)

[4] D. Rutkovskaya, *Neural networks, genetic algorithms and fuzzy systems.* Moscow: Hot line-Telecom, 2006, p. 452. (in Russian)