

Probability Models for Assessing Effectiveness of Advertising Channels in the Internet Environment

Viktoriya I. Tinyakova*

Belgorod State University, Pobedy street 85, 208000 Belgorod, Russia

Valeriy V. Davnis

Belgorod State University, Pobedy street 85, 208000 Belgorod, Russia

Yaroslav B. Lavrinenko

Voronezh State Technical University, Moscow Avenue 14, 394000 Voronezh, Russia

Larisa A. Shishkina

Voronezh State Agrarian University named after Emperor Peter the Great, Michurina street 1, 394087 Voronezh, Russia

Abstract

Nowadays, marketing specialists simultaneously use several channels to attract visitors to websites. There is a difficulty in assessing not only the efficiency and conversion of each channel separately, but also in their interconnection. The problem occurs when users visit a website from several sources and only after that do the key action. To assess the effectiveness and selection of the most optimal channels, different models of attribution are used. The models are reviewed in the article. However, we propose to use multi-channel attribution, which provides an aggregate assessment of multi-channel sequences, taking into account that they are interdependent. The purpose of the paper is to create an attribution model that comprehensively evaluates multi-channel sequences and shows the effect of each channel on the conversion. The presented model of attribution can be based on the theory of graphs or Markov chains. The first method of calculation is more visual, the second (based on Markov chains) allows for work with a large amount of data. As a result, a model of multi-channel attribution was presented, which is based on Markov processes or graph theory. It allows for maximum comprehensive assessing of the impact of each channel on the conversion. On the basis of the two methods, calculations were carried out, confirming the adequacy of the model used for the tasks assigned.

Keywords: Attribution model; Multi-channel attribution; Internet marketing; Advertising channel; Conversion; Advertising budget; Graph model; Markov chain.



CC BY: [Creative Commons Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Recently, successful internet marketers use more and more channels. In addition to search engine optimization and contextual advertising (Yandex.Direct and GoogleAdwords), e-mail channels, social networks, Instagram, remarketing / retargeting, etc. were included in practical use. Therefore, marketers face the task of selecting those advertising channels that will be most effective for a specific project. In addition to the difficulty in choosing the optimal advertising channels, it is difficult to choose a model for the integrated evaluation of the efficiency of a channel for the subsequent distribution of the advertising budget between them.

The first difficulty in complex evaluation is that users can visit the website in several ways: they can go to the site following a direct link, and from social networks, from advertising links on Yandex, etc. Moreover, before making the desired action (conversion)

on the site, users can repeatedly visit the site from different "entry points": the first time they can go to the site by clicking on the advertisement (CPC) from Yandex search request results, the second visit can be by direct link (Direct), and the third one (leading to the conversion - C) can be from a social network (Social) in this case we observe the chain (multichannel sequence):

$$CPC \rightarrow Direct \rightarrow Social \rightarrow C \quad (1)$$

Thus, when evaluating the effectiveness of advertising channels, the marketing specialist first of all needs to answer the questions: how to assess the contribution of a particular source to the formation of conversion on the website? What happens with the conversion on the site, if you exclude this or that marketing channel? To answer this question, there are a number of methodologies, which are called attribution models

2. Methods Literature Review

The attribution model is a way of distributing the "weight" of conversion between channels. Depending on the choice of the attribution model, the weight of the channel (source) will be calculated, which can be conditionally considered as the contribution that this source in conversion. We can review the sequence:

$$AdWords \rightarrow CPC \rightarrow Social \rightarrow E - mail \rightarrow Yandex \rightarrow CPC \rightarrow Direct \rightarrow C \quad (2)$$

The following basic attribution models are distinguished now:

1. By the last interaction - Last Click Model (LCM). This model, for its simplicity and intuitive "correctness", has become most widespread in practice. In the most general case, within the LCM model, all 100% of the weight of the conversion is given to the last channel in the multichannel sequence, which preceded the fact of the occurrence of the desired action.
2. First interaction – First Click Model (FCM). В данной модели 100% вес отдается первому источнику в последовательности и 0% всем остальным.
3. In this model, 100% weight goes to the first source in the sequence and 0% to all the rest.
4. Linear model – Linear Model (LM). Within its framework, all channels get their nonzero weight. In the case of LM, all channels have the same weight (that is, their contributions to the conversion) and are considered to be equivalent.
5. Time Decay – Time Decay Model (TDM). The TDM attribution model is based on the assumption that the contribution of the channel is greater the closer it is to the conversion, so the channel weight is a monotonically increasing function depending on its position in the chain. Based on the position - Position Type Model (PTM). The PTM attribution model is a combination of three models: LCM, FCM and LM. Within its framework, the maximum share (usually 40) goes to the first and last interactions in the chain, and the remaining (typically

20) are distributed evenly (as in the linear model) between the intermediate channels.

The choice of the attribution model is the most important step in assessing the effectiveness of internet marketing (Davnis and Tinyakova, 2006). Depending on the model, the marketing manager can get absolutely opposite conclusions about the profitability of a particular channel. Especially this is observed in the spheres, where there is a long decision-making process, for example, in real estate sphere (Lavrynenko and Tinyakova, 2013). The question arises: which model of attribution should be considered as a reference model?

A number of authors reviewed this problem: Hastie *et al.* (2009), Roberts and Zahay (2012), Li and Kannan (2014).

As a rule, the LCM model is being chosen. However, in practice, there were cases when replacing LCM with PTM and subsequent allocation of funds between channels allowed to significantly increase the efficiency of marketing activities (Hollensen *et al.*, 2017; Kotler *et al.*, 2016; Roos, 2017). The modern models have the following disadvantages:

- the impossibility of obtaining unambiguous results and their evaluation of the choice of a particular model;
- the use of expert choice of the model increases the subjectivity of further decisions;
- combined models are also not deprived of their disadvantages due to the initially selected weights

3. Data and Methodology

The reviewed model was originally developed for aggregate assessment of multi-channel sequences, assuming that the channels are interdependent.

Let's describe the data format which our model interacts with. Assuming that for the analyzed time interval T, M transitions were made to the website, that is, we have data on M user sessions. Each i session S_i has a fixed set of parameters (session attributes) P. For our analysis, we

need the following set of attributes to be included in the set of all session attributes:

$$A = \{SrcType, T, URL, clientID, CVtype\} \in P \quad (3)$$

SrcType – link channel;

URL – the address of the page that the user visited when going to the website;

clientID – the user's unique identifier;

CVtype – whether a conversion was made as a result of the session (CV – yes, N – no);

T – time interval between TimeS; TimeF.

The channel is the source of traffic, which can include: Yandex CPC, Google CPC, Facebook, Vkontakte, Instagram, Direct, Referral, etc..

Advertising channels are coded as follows: c_1, c_2, \dots, c_k assuming that their number is limited by the value of k.

Let's suggest that M sessions $\Sigma = \{S_1, S_2, \dots, S_M\}$ were initiated by $G \leq M$ users. Using the unique identifier of the user clientID we can divide the set Σ into G disjoint subsets:

$$\Sigma = U_1 \cup U_2 \dots \cup U_G \quad (4)$$

U_i multiple sessions (sorted by date ascending order) with the same clientID, i.e., a set of chronologically ordered sessions initiated by the same user. Considering our assumption that $[TimeS; TimeF] \subset T$, then on the basis of data in U_i we can associate with each i user the following chain of channels: where $L_{(i)} = |U_i|$ is the number of elements (the number of the user's transitions to the website) in the set U_i . The above transition chain is a sequence of traffic sources that the i-user used during the interaction with the website.

Next we introduce two additional "pseudo-channels" CV and N according to the rule:

- if during the user's session i with the source $c_{ij} (1 \leq j \leq L_i)$ there was a conversion, then after c_{ij} we add CV, and obtain $\dots \rightarrow c_{ij} \rightarrow CV \rightarrow \dots$;

- if no conversion occurred as a result of the last current session i with the source $C_i L_i$, then after $C_i L_i$ we add N, and obtain $\dots \rightarrow C_i L_i \rightarrow N$.

Let's additionally pay attention to the situation when we have chains like: $\dots \rightarrow c_{ij} \rightarrow CV \rightarrow \dots \rightarrow CV \rightarrow \dots$

Sequences with such a structure cannot arise according to the rules formulated above, but nevertheless they can occur in a number of cases, for example, in call-related topics, where besides the above session parameters we have a unique relation: $clientID \rightarrow TelNumber$.

A key feature of the above method of forming chains of users' interaction with the website is that any chain of interaction (multichannel sequence) always ends as one of two "events": CV or N. Event N can occur only in the end of the sequence, while event CV can occur at an arbitrary place.

Let's perform examples of sequences formed according to the above mentioned rules. For simplicity, we shall take 3 different channels c_1, c_2, c_3 and add CV and N to them:

$$\begin{aligned} c_1 &\rightarrow N; \\ c_1 &\rightarrow c_2 \rightarrow N; \\ c_1 &\rightarrow CV; \\ c_1 &\rightarrow c_2 \rightarrow CV \rightarrow c_1 \rightarrow N; \\ c_1 &\rightarrow c_2 \rightarrow CV \rightarrow c_3 \rightarrow N; \\ c_1 &\rightarrow c_2 \rightarrow CV \rightarrow CV; \\ c_1 &\rightarrow c_2 \rightarrow CV \rightarrow CV \rightarrow c_3 \rightarrow N. \end{aligned}$$

The next step necessary to construct a multi-channel attribution model is to transform the sequences so that event CV, like event N, can occur only strictly at the end of the sequence (such sequences will be called elementary sequences). For this purpose, we shall "split" the original chains so that at their end they will always have a CV or N event.

Let us demonstrate this method by the example of typical sequences:

chains 1-4 modified to "elementary" form;

let us "split" chain 5 into: $c_1 \rightarrow c_2 \rightarrow CV$ and $c_1 \rightarrow c_2 \rightarrow c_1 \rightarrow N$;

let us "split" chain 6 into: $c_1 \rightarrow c_2 \rightarrow CV$ and $c_1 \rightarrow c_2 \rightarrow c_3 \rightarrow N$;

let us "split" chain 7 into: $c_1 \rightarrow c_2 \rightarrow CV$ and $c_1 \rightarrow c_2 \rightarrow CV$;

let us "split" chain 8 into: $c_1 \rightarrow c_2 \rightarrow CV$, $c_1 \rightarrow c_2 \rightarrow CV$ and $c_1 \rightarrow c_2 \rightarrow c_3 \rightarrow N$.

Then we shall review a set of G sequences (we assume that all of them are already elementary, that is, they end in CV or N. We suggest that from X sequences X end with CV and G and X with N. Then we define the effect of channel c_i on conversion on the website for time T through $I(c_i)$, and the elementary chain j through R_j . The value of impact $I(c_i)$ of the channel c_i on conversion will be considered as the number of "lost" conversions in case of removing the channel c_i from all conversion chains, where it is performed, referred to the total number of conversions X:

$$I(c_i) = \frac{|\{R_j | c_i \in R_j, CV=R_j\}|}{X} \quad (5)$$

It is obvious that for any c_i the value of $I(c_i)$ satisfies the following inequality: $0 \leq I(c_i) \leq 1$

Moreover, $I(c_i) = 0$ if and only if channel c_i is not included into any "conversion" sequence, and $I(c_i) = 1$ if and only if the removal of c_i leads to loss of all conversions on the website. Thus, it will be easy to estimate the new number of conversions that will result after removing the channel c_i :

$$CV_{new} = X * (1 - I(c_i)) \quad (6)$$

The sum of the effects of the channels is not equal to one. For convenience, we can introduce a normalization and calculate the normalized influence of

$I_n(c_j)$ channels on conversion:

$$I_n(c_j) = \frac{I(c_j)}{\sum_{i=1}^k I(c_i)} \quad (7)$$

If the task is to find out how the channel c_i affects c_j , then we can use the following argument: the user's session initiated by channel c_i leads to the session with channel c_j as many times as there are chains R_f , where c_i precedes c_j in them. And if we designate the value of such influence as $I(c_i, c_j)$, then:

$$I(c_i, c_j) = \frac{|\{R_f | c_i, c_j \in R_t \text{ and } c_i \text{ precede } c_j\}|}{|\{R_t | c_j \in R_t\}|} \quad (8)$$

Generally, the function $I(c_i, c_j)$ is not symmetric: $I(c_i, c_j) \neq I(c_j, c_i)$. Sequences R_f where c_i precedes c_j and at the same time c_j precedes c_i (i.e. cycles are formed) can also be taken into account in the denominator of the formula. The normalization introduced earlier can be generalized in a natural way to the case described above:

$$I_n(c_i, c_j) = \frac{I(c_i, c_j)}{\sum_{h=1}^k I(c_i, c_h)} \quad (9)$$

3.1. Channel Cost Assessment

To assess the basic metrics, we also need to add to the parameters of user's sessions an indicator called "cost of transition". It can be interpreted as the cost paid by the advertiser, per user click on the given channel, if the channel is free (for example, direct link), then we will assume that the cost of the transition equals 0. If it is possible to determine only the total cost of the channel (for example for SEO), then we assume that the cost of the transition in a

particular session is equal to the ratio of the total costs per channel to the number of uses of this channel for all sessions. We will designate the cost of the transition for channel c_i in chain R_j through $V_j(c_i)$. Thus, we can estimate the cost $V(R_j)$ of one chain R_j as follows:

$$V(R_j) = \sum_{c_i \in R_j} V_j(c_i) \tag{10}$$

Total costs for channel c_i equal:

$$V(c_i) = \sum_{j: c_i \in R_j} V_j(c_i) \tag{11}$$

Total costs of attracting users to the website using channels c_1, c_2, \dots, c_k equal

$$V = \sum_{j=1}^G \sum_{c_i \in R_j} V_j(c_i) = \sum_{i=1}^k \sum_{j: c_i \in R_j} V_j(c_i) \tag{12}$$

The duality of the formula results from different ways of calculating the total costs: in the first case, we sum the costs for each of the chains in all G chains, and in the second case we sum the channel costs for all k channels.

After describing all elements included in models of multichannel attribution, we shall review the two most effective ones: graph and matrix models (Wiesner *et al.*, 2001).

3.2. Graph Model

In order to perform the set of chains in the form of a graph, we need to fix two sets: the set of V vertices and the set of E connections between them. Marketing channels and additional events will be the vertices:

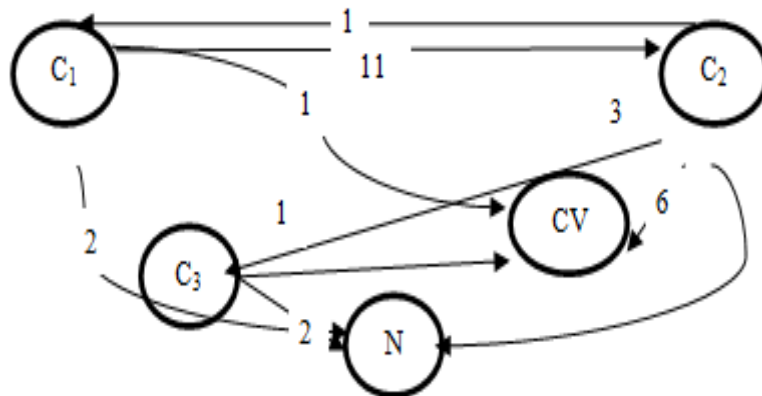
$$V = \{c_1, c_2, \dots, c_k, CV, N\}$$

We shall choose pairs of connected elements of V as E . For the elementary chains reviewed above, we obtain:

$$V = \{c_1, c_2, c_3, CV, N\}, E = \{(c_1, N), (c_1, c_2), (c_1, CV), (c_1, c_2), \dots, (c_3, N)\}.$$

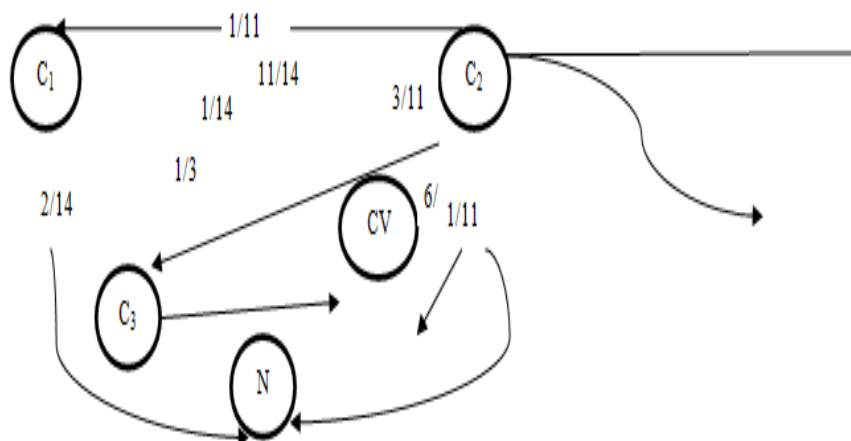
Since there may be coincident elements in the set E , the resulting graph can have multiple (doubled) edges – this will complicate the perception. Therefore, the original graph is transformed into an oriented weighted graph (see Figure 1):

Figure-1. Oriented weighted graph



It should be noted that $P P(c_1, CV)$ is the probability of conversion of source c_1 in the classical LCM model. It is obvious that the LCM model does not take into account the large amount of statistical data that we can collect by analyzing users' sessions. If we perform calculations for all the remaining vertices, then our graph will be transformed to the form below (see Figure 2).

Figure-2. Graph for calculating the total probability of conversion of a specific channel



Based on this model, we can calculate the full probability of conversion for a particular channel. The following recursive formula is used for calculation:

$$P_{full}(c_i, CV) = \sum_{c_j: c_i \rightarrow c_j} P(c_i, c_j) P_{full}(c_j, CV) \tag{13}$$

However, if we assume the possibility of transitions in the graph of type... → c_i → c_i → ... (i.e., permit loops), then the system of equations becomes nonlinear, which considerably complicates determining of the required probabilities.

3.3. Matrix Model

Let us consider the second model of multi-channel attribution – matrix model. Imagine a set of k channels c₁, c₂, ... c_k and two additional "pseudo-channels" CV, N. In the graph model they were performed as vertices. As a result, we can from a square matrix (K + 2) * (k + 2), with conditional probabilities as elements

$$H = \begin{pmatrix} P(c_1, c_1) & P(c_1, c_2) & \dots & P(c_1, c_k) & P(c_1, CV) & P(c_1, N) \\ P(c_2, c_1) & P(c_2, c_2) & \dots & P(c_2, c_k) & P(c_2, CV) & P(c_2, N) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ P(c_k, c_1) & P(c_k, c_2) & \dots & P(c_k, c_k) & P(c_k, CV) & P(c_k, N) \\ 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \tag{14}$$

The matrix for which this condition is satisfied is called stochastic. It is known that an arbitrary stochastic matrix defines a certain random process, called Markov process.

Such models allow to answer a number of important questions, in particular:

- What is the probability of passing from state c_i to state c_j in t steps?
- What will be the distribution of the probability of

finding in each channel in t steps? To solve it, we must find the answer for the special case of the first question: what is the total probability of passing from the state c_i to CV?

$$P_{full}(c_i, CV) = \lim_{t \rightarrow \infty} H^t(i, k + 1) \tag{16}$$

4. Empirical Results

As an example, we shall calculate the total probability of conversion P_{full}(c₁, CV) for source c₁.

Since c₁ is associated with c₂, CV, N, but the probability of passing from N to CV equals zero, and the probability of passing from CV to CV equals 1, then:

$$P_{full}(c_1, CV) = P(c_1, c_1) * P_{full}(c_2, CV) + P(c_1, CV) * 1 = \frac{11}{14} * P_{full}(c_2, CV) + \frac{1}{14}$$

From c₂, we can return to c₁ or pass to c₃, CV, N, which means:

$$P_{full}(c_2, CV) = P(c_2, c_1) * P_{full}(c_1, CV) + P(c_2, c_3) * P_{full}(c_3, CV) + P(c_2, CV) * 1$$

$$P_{full}(c_2, CV) = \frac{1}{11} * P_{full}(c_1, CV) + \frac{3}{11} * P_{full}(c_3, CV) + \frac{6}{11}$$

Next we shall transform:

$$P_{full}(c_1, CV) = \frac{11}{14} * \left(\frac{1}{11} * P_{full}(c_1, CV) + \frac{3}{11} * P_{full}(c_3, CV) + \frac{6}{11} \right) + \frac{1}{14}$$

For convenience, we shall designate P_{full}(c₁, CV) = x., then we obtain the following linear equation:

$$x = \frac{11}{14} * \left(\frac{1}{11} x + \frac{3}{11} * P_{full}(c_3, CV) + \frac{6}{11} \right) + \frac{1}{14}$$

Now we shall calculate P_{full}(c₃, CV). From source c₃, we can only go to CV or N.

As a result we shall obtain:

$$P_{full}(c_3, CV) = \frac{1}{3}$$

Finally we have the following equation:

$$x = \frac{11}{14} * \left(\frac{1}{11} x + \frac{3}{11} * \frac{1}{3} + \frac{6}{11} \right) + \frac{1}{14}$$

And we further we can determine x:

$$x = P_{full}(c_1, CV) = \frac{8}{13} = 0.6154$$

The main advantage of the above model is its clarity, while the obvious disadvantages (as can be seen even in a simple example) include high computational complexity in case of large number of traffic sources.

Let us make the calculation for the matrix model. For the example above, we obtain:

$$H = \begin{pmatrix} 0 & \frac{11}{14} & 0 & \frac{1}{14} & \frac{2}{14} \\ \frac{1}{11} & 0 & \frac{3}{11} & \frac{6}{11} & \frac{1}{11} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

We can prove that for the case when transitions to no other state are possible from states CV and N, this limit exists. Of course, in practice, we cannot operate with the "infinite" degree of the matrix. However, instead of "infinity" it is usually sufficient to take a sufficiently higher power of two. The convenience of raising a matrix to the 2t power is that matrix H has to be multiplied by itself.

Let us show on our example the rate of "convergence" of the limit to the required probability:

H^2	0.0714 0 0 0 0	0 0.0714 0 0 0	0.02143 0 0 0 0	0.5000 0.6429 0.3333 1.0000 0	0.2143 0.2857 0.6667 0 1.0000
H^4	0.0051 0 0 0 0	0 0.0051 0 0 0	0.0153 0 0 0 0	0.6071 0.6888 0.3333 1.0000 0	0.3724 0.3061 0.6667 0 1.0000
H^8	0.0000 0 0 0 0	0 0.0000 0 0 0	0.0001 0 0 0 0	0.6153 0.6923 0.3333 1.0000 0	0.3846 0.3077 0.6667 0 1.0000
H^{16}	0.0000 0 0 0 0	0 0.0000 0 0 0	0.0000 0 0 0 0	0.6154 0.6923 0.3333 1.0000 0	0.3846 0.3077 0.6667 0 1.0000
H^{32}	0.0000 0 0 0 0	0 0.0000 0 0 0	0.0000 0 0 0 0	0.6154 0.6923 0.3333 1.0000 0	0.3846 0.3077 0.6667 0 1.0000
H^{64}	0.0000 0 0 0 0	0 0.0000 0 0 0	0.0000 0 0 0 0	0.6154 0.6923 0.3333 1.0000 0	0.3846 0.3077 0.6667 0 1.0000

As a result, for H^8 the calculated probability $P_{full}(c_1, CV)$ differs from the exact value that we previously obtained on the basis of the graph model, in the fourth decimal place. The probability values calculated for H^{16} , H^{32} , H^{64} coincide. Thus, in this case it was sufficient to limit to calculating H^8 , which requires only 3 matrix multiplications. Thus, the rate of convergence of the limit to the required probability is high enough, which makes this model effective in practical applications.

5. Summary and Conclusion

Currently used classical models of conversion attribution were reviewed. In addition, a multi-channel attribution model based on Markov processes (chains) is described that allows to evaluate comprehensively the probability of conversion for each advertising channel and calculate the impact of the channel on the website conversion. Approaches that allow to adapt the formed model for the optimization of rates in contextual advertising were performed.

References

Davnis, V. and Tinyakova, V. (2006). Modern methods of analysis and forecasting in the problems of justifying marketing decisions. Marketing v rossii i za rubezhom. *Marketing in Russia and Abroad*, 2: 16-26.

- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *Unsupervised learning. In the elements of statistical learning.* Springer: New York, NY. 485-585.
- Hollensen, S., Kotler, P. and Opresnik, M. O. (2017). *Social media marketing, A practitioner guide.*
- Kotler, P., Kartajaya, H. and Setiawan, I. (2016). *Marketing 4.0, Moving from traditional to digital.* John Wiley & Sons.
- Lavrynenko, Y. and Tinyakova, V. (2013). Increase in the conversion of websites as a way to reduce the advertising costs of real estate agencies. *Vestnik saratovskogo gosudarstvennogo sotsialno-ekonomicheskogo universiteta. Bulletin of the Saratov State Social and Economic University*, 2(46): 58-63.
- Li, H. and Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment, An empirical model and a field experiment. *Journal of Marketing Research*, 51(1): 40-56.
- Roberts, M. L. and Zahay, D. (2012). *Internet marketing: Integrating online and offline strategies.* Cengage Learning.
- Roos, J. (2017). *The legal marketing fastlane, Your roadmap to generating real leads in 72 hours or less, even if you're small.* CreateSpace Independent Publishing Platform.
- Wiesner, R. H., McDiarmid, S. V., Kamath, P. S., Edwards, E. B., Malinchoc, M., Kremers, W. K. and Kim, W. R. (2001). Meld and peld, Application of survival models to liver allocation. *Liver Transplantation*, 7(7): 567-80.