МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ В ПРИРОДНИЧИХ НАУКАХ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

DOI:
UDK 339.198
O. Shumeiko[1], Doctor of Technical Sciences, Professor
O. Davidchik[2], Ph.D., Associate Professor
[1]Dnipro State Technical University, Kamianske, Ukraine
[2]Dnipro Vocational College of Engineering and Pedagogy of the State Higher Educational Institution "Ukrainian State University of Chemical Technology", Kamianske

BUILDING OF SALES PROMOTION RATING

The purpose of this paper is to study the task of building rated goods and services based on feedback from customers with subsequent clustering and to identify hidden links between different groups of customers. Our research lies outside the existing methods and is based primarily on feedback between consumers of goods and services and producers.

The development of online advertising makes it necessary to build a rating of goods and services based on customer feedback. The need to deliver the customer to the sector of goods and services that interest him, leads to the task of identifying hidden links between goods and services, which is chosen by a particular group of customers.

Keywords: *rating, online advertising, clustering.*

Метою даного дослідження є вивчення завдання побудови рейтингових товарів та послуг на основі відгуків від клієнтів з подальшою кластеризацією та виявлення прихованих зв'язків між різними групами клієнтів. Наші дослідження лежать осторонь існуючих методів і засновані насамперед на зворотному зв'язку між споживачами товарів та послуг та виробниками.

Розвиток інтернет-реклами набуває необхідності побудови рейтингу товарів та послуг на основі відгуків від клієнтів. Необхідність доставки клієнта в той сектор товарів та послуг, який йому цікавий, призводить до завдання виявлення прихованих зв'язків між товарами та послугами, який обирається тією чи іншою групою клієнтів.

Ключові слова: *рейтинг, інтернет-реклама, кластеризація.*

Problem's Formulation

The purpose of this paper is to study the problem of constructing a rating of goods and services based on reviews from customers, followed by clustering and identifying hidden connections between different groups of customers. Methods for researching the effectiveness of advertising based on the behavioral function of clients have been carried out for a long time. Traditionally, they are divided into methods of collaboration filtering and methods of content analysis. These methods are briefly reviewed in [1].

Our research is outside the existing methods and is based primarily on feedback between consumers of goods and services and manufacturers.

The coronavirus pandemic has led to a significant change in the relationship between customers (consumers of services and goods) and, accordingly, their manufacturers.

In these conditions, Internet advertising is gaining a great role. Restricting the movement of people leads to the fact that the role of the Internet is increasing, both in terms of information and commutative connections, and in terms of the sale of goods and services. The development of online advertising is acquiring the need to build a rating of goods and services based on customer reviews. Customer opinion has become more decisive than ever. The need to deliver a client to the sector of goods and services that is of interest to him leads to the task of revealing the hidden links between goods and services, chosen by a particular group of clients.

Formulation of the study purpose

The research data is based on the method of statistical analysis. First of all, we point out that the following predicating states.

Theorem (*the binomial distribution*). Let us assume that p is an A event rate. Thereafter the probability, that from n independent trials there will be equally k favorable results, is equal to

$$P_n(k) = C_n^k p^k (1-p)^{n-k},$$

that is Bernoulli distribution.

The binomial distribution is one of the most frequently encountered discontinuous distributions that are used as a chance model of many phenomena. It occurs in case when we need to know how many times some event happens from present amongst of independent observations (trials) that are carried out in the same conditions.

Let us consider any mass production. Even in its normal operation, sometimes the products which do not conform to the standard, i.e. defective, are made. Denote the deal of the defective products with p, $0 < p < 1$. It is impossible to tell preliminarily (before its production) what product exactly will be found unqualified. The following symbolic model is usually used to describe such a situation:

1) each product with a probability of p can be found defective (with a probability $q = 1 - p$ it conforms to the standard); this probability is the same for all the products;

2) defective as well as normal products appear independently of one another. This means that in normal manufacture process the appearance of unqualified product does not influence the possibility of the defective product appearance in future. The violation of this rule means the malfunctioning of the processing method.

The consecutive order of the independent trials, where the result of each can be one of two outcomes (for example, success or failure) and the probability of the «success» (or «failure») in each trial is the same, is called the Bernoulli test pattern. As a consequence, we can restate mentioned above as: in normal conditions the technological process of production mathematically is presented by the Bernoulli test pattern.

Subsequently we need β-distribution

$$g(p) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \quad B(\alpha,\beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp.$$

The β-distribution represents the probability distribution of probabilities, that is to say, as in this case the probability is shaped, and then the range of definition represents the interval [0, 1]. It is not hard to see the resemblance of the β-distribution and the Bernoulli distribution, but the difference between them is that the Bernoulli distribution shapes the event quantity k and the β-distribution shapes the probability p. By this means in the binomial distribution the probability is a parameter, but in the β-distribution is an accidental variable.

In the binomial distribution indexes of power align with the number of favorable and unsuccessful results, for that reason it is naturally to consider that for the β-distribution $\alpha - 1$ coheres the number of favorable results and $\beta - 1$ the number of unsuccessful ones, same as k and $n-k$ in the binomial distribution.

Let us assume after watching the advertisement we have the following statistical law of the advertising account site attendance

$$\mathrm{B}(300,200)=\frac{1}{B(300,200)}p^{300}(1-p)^{200},$$

it means $\alpha-1=300, \beta-1=200$.

Let us find the minimal fiducial advertisement characteristic, number X, that shows the minimal appraisal which the advertisement can receive after the0 infinite number of broadcastings. We will consider 90 % as a level of confidence. This means that we want to be 90 % confident that the advertisement will work not worse than X, so

$$\int_X^1 \frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}=0{,}9.$$

Diagrammatically it means that 90 % of space under the diagram has to be to the right of X (fig. 1).

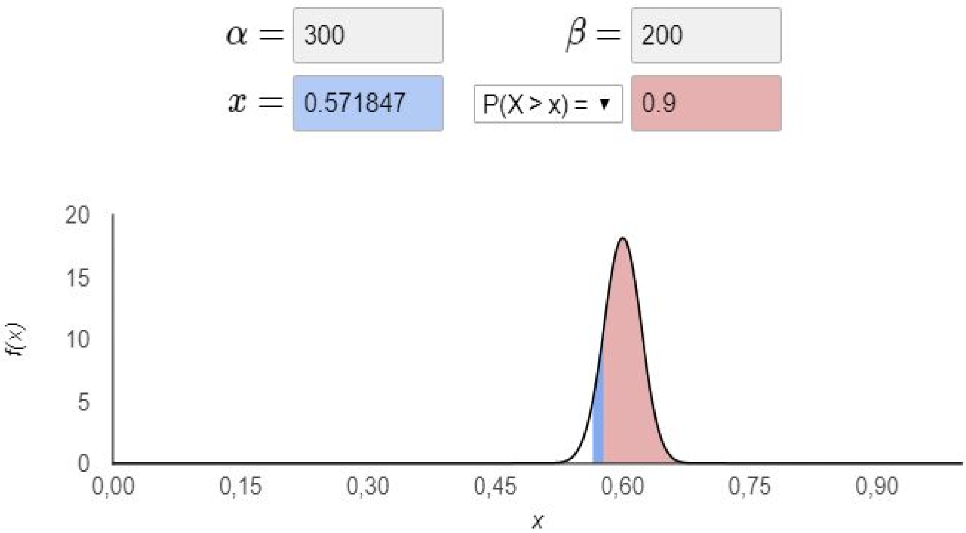

Fig. 1

In our case X = 0.571847. So, with accuracy to 90 % we get that 0.571847 of all who saw the advertisement would go to the advertising account site.

Let us consider the following matter — some amount of people saw the previous advertisement and some part of them reacted on it somehow, someone went to the advertising account site and someone ignored it. It is required to make a forecast of the effectiveness of the next advertising action. Binary data 0 is the advertisement is ignored, 1 is the advertisement led to an advertising account site visiting.

The Bayesian approach consists of the following steps:

1. The *prior* probability. We choose the probability density to make a model of reference prior distribution $P(\theta)$. It is the best supposal of the parameters before we get the data *X*.

2. We choose the probability density for $P(X\mid\theta)$, which means we make a model of the *X* behavior with an aimed parameter *θ*.

3. The *posterior* probability. We find the posteriori distribution $P(\theta\mid \mathrm{X})$ and choose *θ* with the highest $P(\theta\mid \mathrm{X})$.

Presenting main material

As a result, the posteriori distribution becomes a new prior distribution. We should repeat the third step each time we get new data. Now in more detail, do step-by-step.

Step 1.The prior probability $P(\theta)$.

The goal is to choose the probability density to make a model of θ parameter. The θ parameter shows the probability of the advertisement success. For this purpose we use the β-distribution. In our case α is the number of successful outcomes and β is the number of people who did not react on the advertisement (the number of failure).

Let us assume that from 100 people who saw the advertisement, 20 went to the advertising account site (fig. 2).

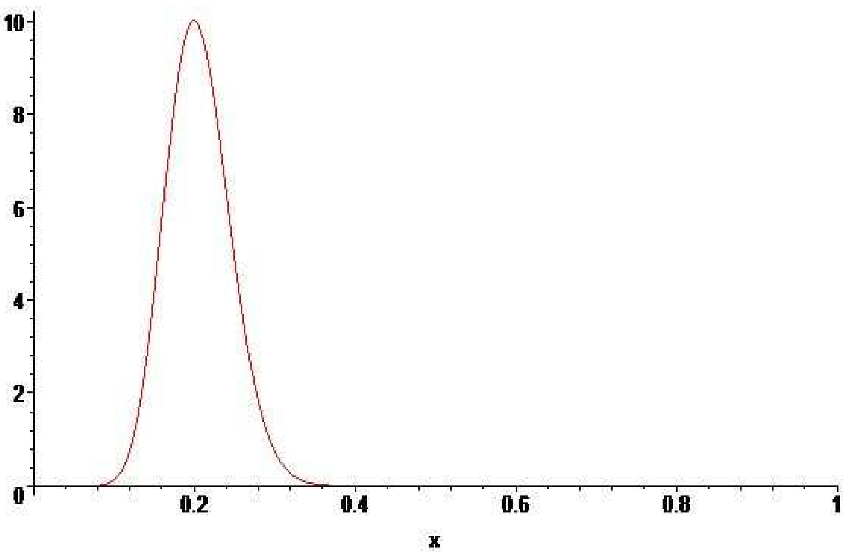

Fig. 2

As we see, moral expectation $\mu = E(X) = 0,2$ of the total number, that means we get not inconsiderable prior probability.

Step 2. The probability function $P(X \mid \theta)$

We choose the probability model for $P(X \mid \theta)$ is the probability of watching the data X is aimed by the θ parameter. The probability function is also called the sampling distribution.

$X = \{x_i\}_{i=1}^{n}$ is a binary set of lattice points [0,0,1,0,1,...,1,0,0,1,1]. For this random variable the distribution density of the probability is equal to

$$f(x_i,\theta) = \theta^{x_i}(1-\theta)^{1-x_i}, \quad x_i = 0,1.$$

In this case the likelihood function is represented as

$$l(\theta,x) = \theta^{s}(1-\theta)^{n-s}, \quad \theta \in (0,1),$$

where $s = \sum_{i=1}^{n} x_i$ is the total number of successful results in n trials, fortuitous of the variable s is conversely the realization of the random variable which is distributed along the binomial law.

Then

$$\frac{d}{dx} l(\theta,x) = \frac{s}{\theta} - \frac{n-s}{1-\theta} = \frac{s-n\theta}{\theta(1-\theta)}$$

and the maximum is attained when $\hat{\theta} = \frac{s}{n}$, which equals to a fraction of successful outcomes in the set of n trials.

Then if in the experiment, which is observed, of 55 people who saw the advertisement only 5 reacted on it, we get (fig.3).

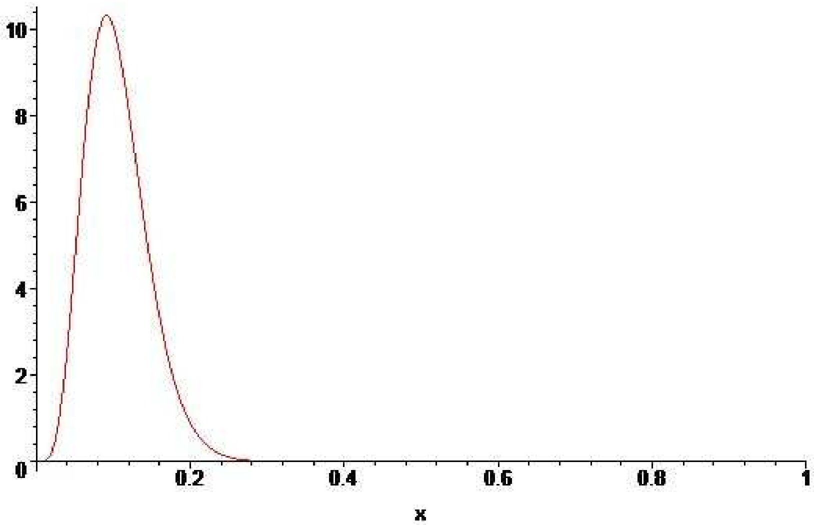

Fig. 3

Step 3. The posterior probability $P(\theta \mid \mathrm{X})$

Our starting supposal about the parameters was $P(\theta)$. Once available the new data, we modify $P(\theta)$ into something more informational namely $P(\theta \mid \mathrm{X})$, relying not only on $P(\theta)$, but also on the present-day data namely $P(X \mid \theta)$, for this purpose, in accordance with the Bayesian formula, we count $P(\theta)$ and $P(X \mid \theta)$ for the prescribed value θ and multiply them. If we make it with each possible θ, we can choose the highest $P(\theta) \cdot P(X \mid \theta)$ among the various θ (fig. 4).

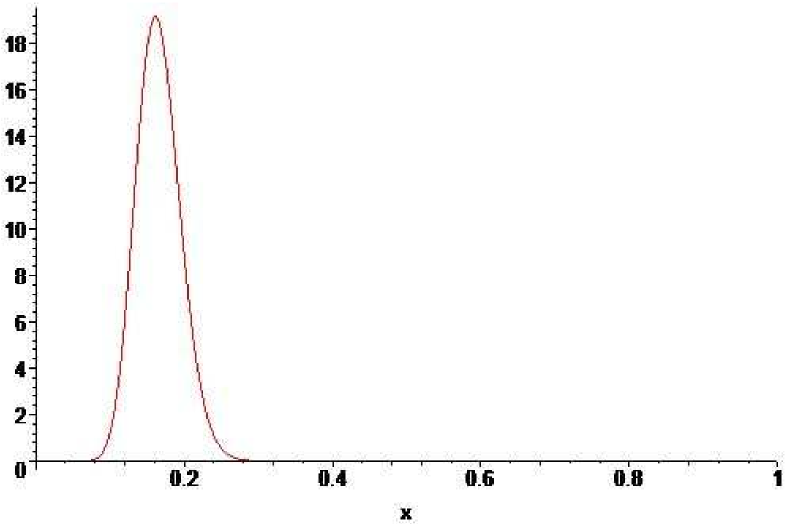

Fig. 4

Please note that in this case the probability shifted to prior. The probability of success in the prior distribution was 20 % and for the aimed data was 9 %. The peak of the posterior probability is approximately 16 %. Besides, the width of the "bell" in the prior probability and the probability function declined in the posterior probability. The spectral band of admissible parameters became narrower, because we use more information in the selection process. The more data we have, the more the posterior probability graph will look like the probability function graph, but not like the prior probability one. In other words, the more information we have, the less the primary prior distribution means.

We show that if we use the β-distribution as a prior distribution then the posterior distribution of the binomial likelihood is also the β-distribution.

Let us assume that

θ is success probability,

x is number of the successful outcomes,

n is total number of trials, it means that $(n-x)$ the number of failures.

We write down the Bayesian formula

$$P(\theta \mid x)=\frac{P(\theta \mid x)P(\theta)}{\int P(\theta \mid x)P(\theta)d\theta}=\frac{C_x^n\theta^x(1-\theta)^{n-x}\frac{1}{B(\alpha,\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 C_x^n\theta^x(1-\theta)^{n-x}\frac{1}{B(\alpha,\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta}=$$

$$=\frac{\frac{C_x^n}{B(\alpha,\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{\frac{C_x^n}{B(\alpha,\beta)}\int_0^1\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}d\theta}=\mathrm{B}(x+\alpha,n-x+\beta).$$

The prior distribution $P(\theta)$ coheres with $\mathrm{B}(\alpha,\beta)$. After getting x successful outcomes and $n-x$ failures in experiments, the posterior distribution also becomes the β-distribution with $(x+\alpha,n-x+\beta)$ parameters.

As the result, the β-distribution is the conjugate prior distribution of the binomial distribution. What does it mean? We already know at the stage of modeling that the posterior distribution will also be the β-distribution. Consequently, after the bigger amount of trials we can count the posterior probability by simple supplementing of number of successes and failures and current parameters α,β, respectively, instead of multiplying the probability function and prior distribution.

Our next purpose is the *clustering of the distribution function*. Let us consider the following matter. Let us assume that there are multitude distribution functions $\Gamma=\{g_i(p)\}, p\in[0,1]$, where $i=0,1,...,n$ the multitude of advertisements that where broadcasted earlier. For the advertisement (trial) $g_i(p)$ we need to forecast the result of the advertisement (trial). It is as if the result is directly relating to the type of the distribution function, this means that we need to cluster the multitude Γ from the clause of distribution functions resembling, find the cluster that is the most suitable for $g_i(p)$ and link the center of this cluster with the forecast of the advertisement.

In case if (fig. 5).

$$g_i(p)=\frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(-\frac{(p-\mu_i)^2}{2\sigma_i^2}\right)$$

all the results are described with normal distribution, then we can put a point (μ_i,σ_i) respectively to each function $g_i(p)$ and make a clustering according to the multitude of points that where got.

If we have the β-distribution, at first view, we can put a point (α,β) respectively to each function and the less the distribution functions will differ, the closer respective points will be.

That is good, but incorrect, for example, $B(1,101)$ it looks as following (fig. 6) and $B(2,101)$ (fig. 7).

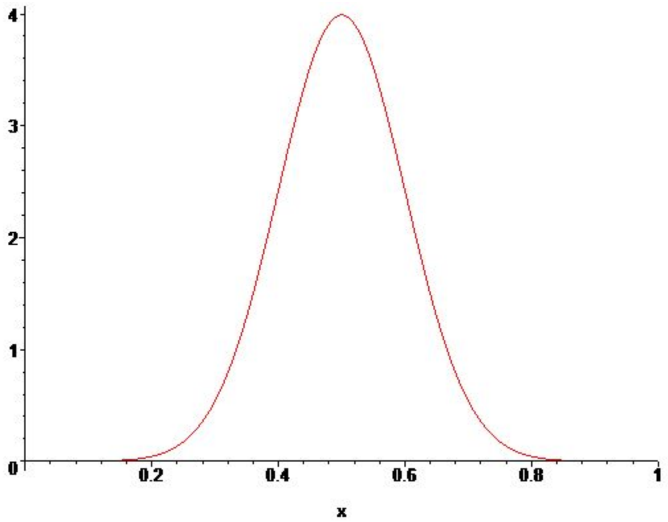

Fig. 5

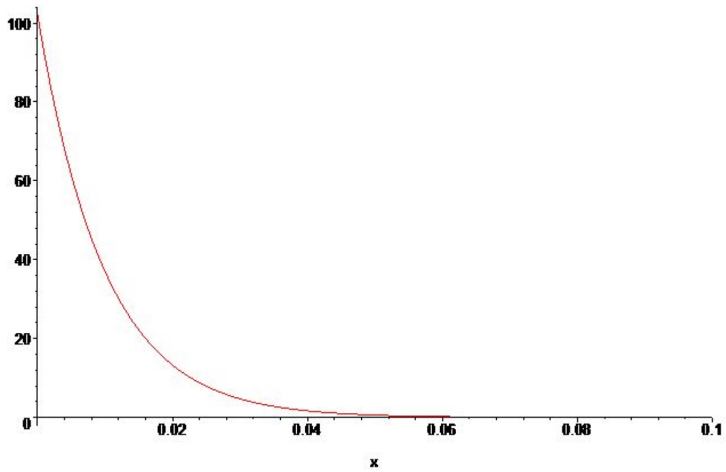

Fig. 6

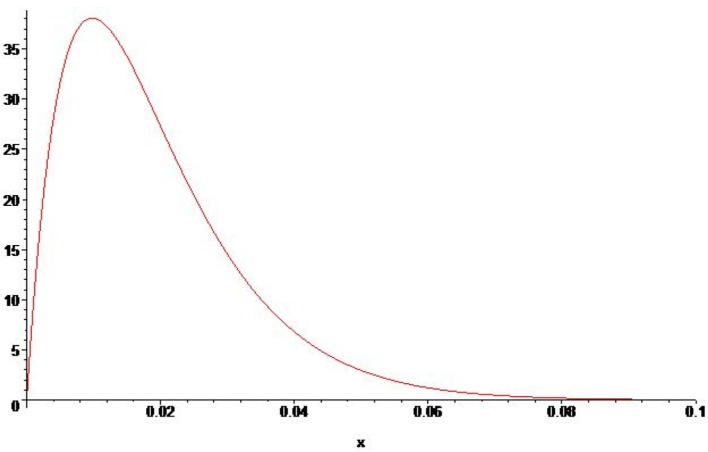

Fig. 7

The points (1,101) and (2,101) differ slightly, but the distribution functions look absolutely different.

Although, if α and β are large enough and do not differ a lot, then this β-distribution can be approximated with normal distribution with

$$\mu = \frac{\alpha}{\alpha+\beta}; \sigma = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} .$$

As in this example (fig.8) $\alpha = 100, \beta = 500$ (the graph of the β-distribution is red), then $\mu = 0{,}16(6); \sigma = 0{,}01520185255$ (the graph of the Gaussian function is green).

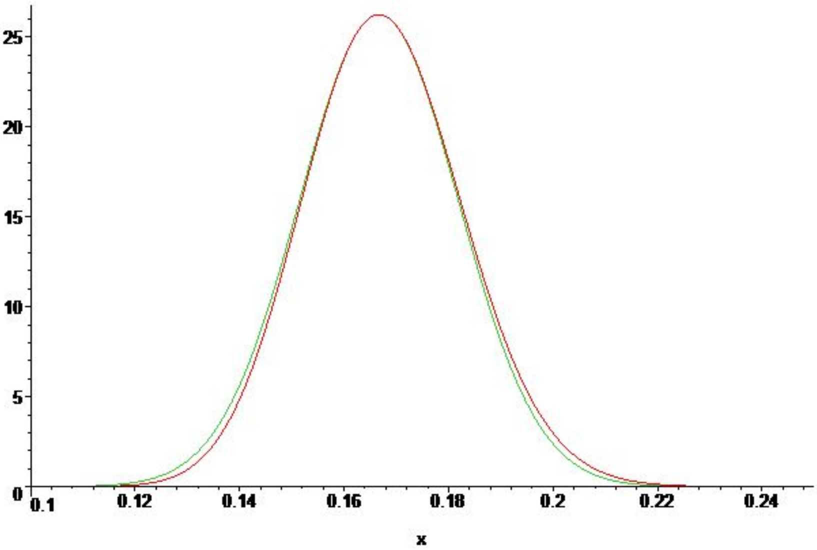

Fig. 8

But then again generally it is incorrect.

In this example (fig. 9) $\alpha = 2, \beta = 500$ (the graph of the β-distribution is red) then $\mu = 0{,}0039840637450; \sigma = 0.002808744857$ (the graph of the Gaussian function is green).

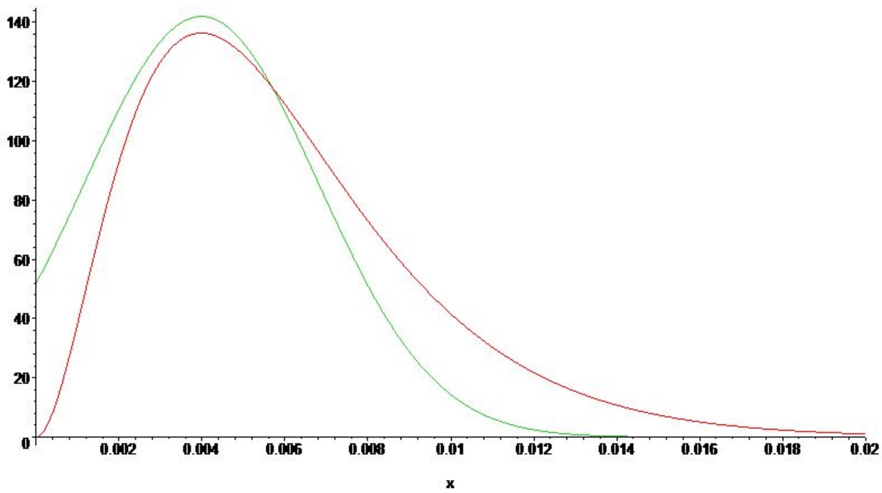

Fig. 9

Let us consider the alternate approach to clustering of the integrated functions $\{f_i(x)\}_{i=0}^{N}$ (in particular, the distribution functions), that are reassigned on the same interval, let it be [0,1].

Let us take the L_1 distance as a difference between two functions $f_i(x)$ and $f_j(x)$

$$\varepsilon_{i,j} = \int_0^1 \left| f_i(x) - f_j(x) \right| dx$$

Here we get the matrix of functions differences (tabl. 1)

Table 1

	$f_0(x)$	$f_1(x)$	$f_2(x)$	…	$f_N(x)$
$f_0(x)$	0	$\varepsilon_{0,1}$	$\varepsilon_{0,2}$	…	$\varepsilon_{0,N}$
$f_1(x)$	$\varepsilon_{1,0}$	0	$\varepsilon_{1,2}$	…	$\varepsilon_{1,N}$
$f_2(x)$	$\varepsilon_{2,0}$	$\varepsilon_{2,1}$	0		$\varepsilon_{2,N}$
…	…	…	…	…	…
$f_N(x)$	$\varepsilon_{N,0}$	$\varepsilon_{N,1}$	$\varepsilon_{N,2}$	…	0

and put

$$\varepsilon = \begin{pmatrix} 0 & \varepsilon_{0,1} & \varepsilon_{0,2} & \dots & \varepsilon_{0,N} \\ \varepsilon_{1,0} & 0 & \varepsilon_{1,2} & \dots & \varepsilon_{1,N} \\ \varepsilon_{2,0} & \varepsilon_{2,1} & 0 & \dots & \varepsilon_{2,N} \\ \dots & \dots & \dots & \dots & \dots \\ \varepsilon_{N,0} & \varepsilon_{N,1} & \varepsilon_{N,2} & \dots & 0 \end{pmatrix}.$$

We use the method of metric scaling to the received data. For this purpose at first we get twice aligned matrix, so we bring to the form which has a mean observation of the numbers in any line and column equal to zero. Twice aligned matrix is uniquely calculated from the original.

At first we build an intermediary matrix ordered $(N+1)\times(N+1)$:

$$J = I - \frac{1}{N+1}\cdot 1 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots. & \dots & \dots. \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} - \frac{1}{N+1}\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}.$$

We mark the matrix from the squares of given data with P

$$P = \begin{pmatrix} 0 & \varepsilon_{0,1}{}^2 & \varepsilon_{0,2}{}^2 & \dots & \varepsilon_{0,N}{}^2 \\ \varepsilon_{1,0}{}^2 & 0 & \varepsilon_{1,2}{}^2 & \dots & \varepsilon_{1,N}{}^2 \\ \varepsilon_{2,0}{}^2 & \varepsilon_{2,1}{}^2 & 0 & \dots & \varepsilon_{2,N}{}^2 \\ \dots & \dots & \dots & \dots & \dots \\ \varepsilon_{N,0}{}^2 & \varepsilon_{N,1}{}^2 & \varepsilon_{N,2}{}^2 & \dots & 0 \end{pmatrix}.$$

Then matrix

$$B = -\frac{1}{2} J \cdot P \cdot J$$

will be twice aligned. Here we get eigenvalues $\lambda_0 \geq \lambda_1 \geq ... \geq \lambda_N$ and eigenvectors $e_0, e_1, ..., e_N$.

We choose $k << N$ the highest eigenvalues $\lambda_0 \geq \lambda_1 \geq ... \geq \lambda_{k-1}$ and respective eigenvectors $e_0, e_1, ..., e_{k-1}$.

Then the projection of data on k principal components will be equal to

$$E \times \sqrt{\Lambda} = \begin{pmatrix} e_0 & e_1 & ... & e_{k-1} \end{pmatrix} \times \begin{pmatrix} \sqrt{\lambda_0} & 0 & ... & 0 \\ 0 & \sqrt{\lambda_1} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \sqrt{\lambda_{k-1}} \end{pmatrix},$$

this means that the dot in *k*-dimensional space will be aligned with each function $f_i(x)$ (with $k = 2$ is a subspace). Using the traditional methods of clustering to the multitude of dots that were gained we get the solution of required matter.

Conclusions

On the basis of the offered algorithm of the Internet advertisement quality estimation, we made a program realization which showed us the effectiveness of the constructed method. The following facts were also established:

1) Direct use of the β-distribution to analyze the effectiveness of advertising is not always justified. In extreme cases, it leads to distortion of the results of the effectiveness of advertising.

2) The task of clustering research results based on approximating the β-distribution by a normal distribution is effective in the case of good customer reviews and ineffective in the presence of a large number of negative reviews, which makes this approach incorrect. Producers of goods and services are receiving incorrect information.

3) The use of clustering based on the β-distribution, given on the same interval, allows us to get around this problem. The use of the metric scaling method allows for effective clustering of online advertising based on customer feedback.

References

1. **Petrov, A**. Introduction to Data Mining/ A. Petrov, O. Shumeiko, B. Basiura .− Krarow: AGH, 2019. − 364 p.
2. **Davison, M.** Multidimensional scaling. Methods of pictorial presentation of data / M. Davison.− M.: Finance and statistics, 1988. − 254 p.
3. **Ayvazyan, S.A.** Practical statistics. Classification and dimensionality reduction/ S.A. Ayvazyan, V.M. Buhshtaber, I.S. Yenyukov, L.D. Meshalkin.− M.: Finance and statistics, 1989. − 607 p.
4. **Esbensen, K.** Multivariate data analysis/K. Esbensen.− Chernogolovka: IPCP RAS, 2005.−160 p.

ПОБУДОВА РЕЙТИНГУ РЕКЛАМНИХ АКЦІЙ

Шумейко О.О., Давидчик О.М.

Реферат

Метою даного дослідження є вивчення завдання побудови рейтинга товарів та послуг на основі відгуків від клієнтів з подальшою кластеризацією та виявлення прихованих зв'язків між різними групами клієнтів. Методи дослідження ефективності реклами, що ґрунтуються на по-

ведінковій функції клієнтів, проводилися досить давно. Традиційно, вони поділяються на методи колабораційної фільтрації та методи контентного аналізу. Наші дослідження лежать осторонь існуючих методів і засновані насамперед на зворотному зв'язку між споживачами товарів та послуг та виробниками.

Пандемія коронавірусу призвела до суттєвої зміни відносин між клієнтами (споживачами послуг та товарів) та, відповідно, їх виробниками. У цих умовах велику роль набувають інтернет-реклама. Обмеження переміщення людей призводить до того, що зростає роль інтернету, як з погляду інформаційно-комутативних зв'язків, і з погляду продажу товарів та послуг. Розвиток інтернет-реклами набуває необхідності побудови рейтингу товарів та послуг на основі відгуків від клієнтів. Думка клієнтів стала, як ніколи, визначальною. Необхідність доставки клієнта в той сектор товарів та послуг, який йому цікавий, призводить до завдання виявлення прихованих зв'язків між товарами та послугами, який обирається тією чи іншою групою клієнтів.

Були встановлені наступні факти:

1. Пряме використання β-розподілу для аналізу ефективності реклами не завжди є виправданим. У крайніх випадках вона призводить до викривлення результатів ефективності реклами.

2. Завдання кластеризації результатів дослідження, заснованих на апроксимації β-розподілу нормальним розподілом, ефективне у разі хороших відгуків клієнтів і неефективне за наявності великої кількості негативних відгуків, що робить цей підхід некоректним. Виробники товарів та послуг отримують неправильну інформацію.

3. Використання кластеризації на основі функції β-розподілу, заданих на тому самому проміжку, дозволяє обійти цю проблему. Застосування методу метричного шкалювання дозволяє провести ефективну кластеризацію інтернет-реклами на основі відгуків клієнтів.