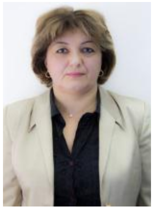
Review of Information Engineering and Applications
2018 Vol. 5, No. 2, pp. 41-48
ISSN(e): 2409-6539
ISSN(p): 2412-3676
DOI: 10.18488/journal.79.2018.52.41.48

CREATION OF ALGORITMS FOR RECOMMENDATION SYSTEM BASED ON USERS DATA ON INTERNET ADVERTISEMENT MARKETING

Kamala K.Hashimova[1]

[1]*Institute of Information Technology ANAS, Baku, Azerbaijan*
Email: multimedia@iit.science.az



Article History
Received: 13 August 2018
Revised: 17 September 2018
Accepted: 23 October 2018
Published: 19 November 2018

Keywords
Internet-advertising data
Hybrid recommendation systems
Filtration
Collaboration filter
Proximity measure.

ABSTRACT

Nowadays, products are offered to buyers in different varieties and qualities on the Internet environment. Recommedation systems are needed to make the right choices and make effective decisions. The article offers a new method and algorithm based on data obtaned through different ways and by creating hybrid recommendation systems. Estimation method is given based on information about objects and users for proper development of the algorithm. The proposed method can identify the proximity between the users group and the objects the users are interested in.

1. INTRODUCTION

In modern days, majority of transaction operations are realized by using digital commersion methods over the internet. Person who wants to obtain product looks over suitable internet shop sites and advertising product sites [1, 2]. Taking a glance at advertising product sites, it is possible to encounter names of different products and services which have analogical functions and identification, but the different prices and characteristic. Lack of information about interesting product, outnumber variety of offerred products on the internet environment lead to wrong decision-making. In the modern times, social networks have spesific roles for recommedation. Improving the digital trade, advertising sites, searching systems make easy for obtaining information about product and services, simultaneously, dynamic change of data makes users select the product [3]. Another interesting fact is a consumer, who offers products and services, sometimes acts as a buyer. Taking into account the current situation, the dynamic changes and right decision-making become difficult for companies in the market. They try to improve by learning their mistakes [4]. On the Internet, recommendation systems are widely used for helping customers and users. Recommendation systems are software tools for recommendation, finding, identifying service, products, objects which may be interesting for users [2]. Then, information is identified taking into account its relevance and importance according to the user's interests. Multidisciplinary knowledge of people is used during creation of recommendation systems. Artificial intelligent, interactions of human and computers, information technologies, data mining, statistics, adaptive user interfaces, making decision systems, marketing and other fields are included to these areas [2]. Recommendation systems are created based on two strategies.

1. Filer operation according to content

2. Collaborative filter operation

Here profiles of users and objects which may be interested them are created. Users can enter demographic data to their profiles. Profiles of objects reflect different attributes depending on their types. Recently, recommendation systems are applied to areas which are not specific for them. For example, the application of recommendation systems in solution of diagnostic issues of complex technical systems [5].

2. PROBLEM STATEMENT

Hybrid recommendation systems increase the effectives of recommendation systems by combining advantages of collaborative algorithms and content-based algorithms. While the complexity of the system increases, the punctuality of the potential recommendations may increase. In the case of information lack, the advantage of the hybrid method may assert itself for collective filtering. When applying the hybrid method, firstly the content is measured on the basis of content filtering and then from these dimensions, mixture of collective filtering is created. As a result, a data set can be formed on the interests of a specific user's activity [6-9].

Recommendation systems can use different algorithms. Obtained results may change depending on the specific issue and the relations of data set.

Regardless of the use of any type of method or algorithm, the recommendation systems use the following considerations in the recommended elements:

- Individual approach - analysis of a certain user's profile, analysis of the relationships in the known conditions
- Social-Collaborative approach - the profiles of other users that can affect the user's choice, the analysis of the advantages and the value they give to specific things
- Content analysis on the objects offered to the users
- Gaining trust by confirming the quality of the offered objects to the users

In practice, it can be observed that a user evaluates the objects differently rather than any other object. Therefore, this evaluation should be considered more informative. If any given evaluation of object is too low or too high than the average evaluation for these objects, these products will be suspicious. A sharp difference in prices should be explained seriously. In that case, recommendation systems should ensure decision making based on the interests of the users.

The purpose of this article is solving the following issues:

- Developing the method for the determination of a group of similar users and close users, who are close to one another due to a specific group of signs by using information obtained from profiles of users and objects and from other sources
- Developing an algorithm for listing the objects that can be of interest for any group of people or for a specific person based on obtained information.

3. PROBLEM SOLUTION

When designing the recommendation system, first of all, it is necessary to pay attention to the evaluation of the results. There are different criteria for evaluating the results. Innovation, accuracy, surprising possibility, robustness to external forces, persistence, and etc. are included here. The accuracy criterion is widely used in practice and it shows how close the given predictions are to the results which may be accepted as an etalon.

After determining the evaluating criteria of results, users' evaluation for different objects and things are studied. The line and column vectors of the future value matrix are generated from the values of other users for services and things. At this time, it should be taken into consideration that offering same products or services by various companies, manufacturers are referred to as different objects. On the other hand, the recommendation system is not created in general, but in a specific area. These systems are created for enterprises selection in the field of education; attractive tourism facilities, routes in tourism; selection of services and prices in healthcare. Different directional and characteristic recommendation systems can be created in accordance with each field. In user profiles, different sides of any object may be evaluated in different aspects. It is possible to create a set of values by a user's values for various objects depending on objects selection in different sources.

Users can respond to different questions during visiting sites. Given values about the usefulness of objects and importance can be used without additional operations over the given values. These types of values can be saved by calculating as average value related on specific time sequences. An average value and the number of the elements generating this value can be saved as a two-dimensional massive in accordance with the time interval. In this case, the two-dimensional time sequence can be written as:

$$\{(k_1, a_1),(k_2, a_2),\ldots,(k_i, a_i)\ldots\} \tag{1}$$

Here a_i is an average value in the i^{th} time interval, k_i- the number of elements generating the average value in the i^{th} time interval

Then $\{(k_1, a_1),(k_2, a_2),\ldots,(k_i, a_i)\ldots\}$. It is important to save the values in the order of time (1) as at any time interval. If the recommendation system encounters any value that related to any previous time interval in practice, then it can recalculate the number and average value at that time interval. If the newly found value is included in the i^{th} time interval and value is b_i, then a_i and k_i can be calculated again.

$$a_i = \frac{k_i * a_i + b_i}{k_i + 1}., \qquad k_i = k_i + 1. \tag{2}$$

The texts that express the attitude of the user to objects can be divided into different parts at each time interval. It may be important to divide a user's comments due to objects and form of value for each object. The methods which are applied for forming user's evaluation in the articles can be used in the entry of recommendation systems, avoiding errors during user enter name for receiving advice on information about objects, things, services. The system can choose the word or phrase combination that is most closely related to its keyword, comparing it with its keywords in its library. If necessary, the user can be required to confirm the found word. As a result, it can be searched according to found word and can generate recommendation. For this purpose, two dimensions, two sets, or two vector proximity can be used.

In simple cases, Jaccar proximity can be used. This measurement is known proximity measures in science; however it is widely used in informatics, molecular biology, ecology, bioinformatics and other areas. When evaluating the proximity of two A and B words in Jaccar meaning:

$$sim(A,B) = \frac{n(A \cap B)}{n(A) + n(B) - n(A \cap B)} = \frac{n(A \cap B)}{n(A \cup B)} \tag{3}$$

At this time, if the difference measure have to be calculated, then it can be written as follows:

$$dif(A,B) = 1 - \frac{n(A \cap B)}{n(A) + n(B) - n(A \cap B)} = \frac{n(A \cup B) - n(A \cap B)}{n(A \cup B)} - \tag{4}$$

It is shown that $0 \le sim(A,B) \le 1$, in special case,

$$sim(A,B) = \begin{cases} 1, A = B \\ 0, A \cap B = \emptyset \end{cases}$$

Due to shown condition, product may be in several categories.

4. ALGORITHM OF PRODUCT EVALUATION

Let's review the proximity of two sentences or two articles instead of two words. Our goal is to generate possible numerical values from the user's object description of the text. There are methods and algorithms, which are used with numerous words and phrases for solving analogical issues in literature [10]. The comments posted by the user on the webpage, messages sent by SMS, or textual responses to the inquiries are small in size. Therefore, these texts can be stored and processed with different algorithms.

First of all, the alphabet should be created to form the evaluation. For example, the weight vector (p_1, p_2.....p_k) for positive attitudes can create relation alphabet by using these words - "excellent", "good", "useful", "advantageous", "sufficient" or for negative attitude, it is possible to create relation alphabet with internal weight vector (m_1,m_2... m_k) by using "worst", "bad", "useless", "disadvantageous", "harmful". Another alphabet used in practice can also be applied. The effectiveness of this alphabet created during experimental experiments can be determined and corrected. In order to correct the alphabet, some ideas of users about objects can be analyzed by experienced experts.

Thus, the value given by the user to the object from the text source (Txt array) can be formed by the following algorithms:

Step 1: Writing the text to the Txt array separated by the natural separators, determining the size of the array and writing to the integer-type T_c parameter [11].

Step 2: Every element in the alphabet is compared to the words in the Txt array, and Jaccar's proximity measure is determined by using formula (4). For this purpose, the dimensions array is reset beforehand. Then:

$$\forall j \in [1, N_i], \quad T(j) = 0$$

$$\forall j \in [1, N_i] \text{ for}$$

$$T(j) = \max\{sim(L_j, Txt_1), sim(L_j, Txt_2), \ldots, sim(L_j, Txt_{Tc})\}$$

Here, L_j is the j-th element of the vector L consisting of the string-type client relationship alphabet. Each alphabet is compared to the customer's comments (Txt_i)

The final value will be $m_q = max\{T(j),\ j \in [1, N_l]\}$

This parameter is a numeric expression of the customer's value given to the objects in the form of text, and may include the following values:

$m_q \in [0, 1]$.

By scaling, this value range can be extended to any other range. If the value m_q is required to be brought to the range $[-a, a]$, then

$m_q = 2a * m_q - a = a * (2m_q - 1)$.

After calculation of this value, it can be added to the time sequence (3) by the formula (2) in accordance with the writing date of the texts.

It is known that the recommendation systems form the current value, taking into account the customer's previous values, and this current value is included in the user value array. The issue of forming a current value from the time sequences value may be calculated as follows:

1. Smoothing the obtained time sequence. For this purpose, different smoothing methods can be used for time series
2. Determining application of the current value with autoregressive methods. In simple case, the weighted moving average value method can be used for this purpose:

(2) taking *k* for the depth of historical memory by using time sequence and taking current value as r_{ij} then (the current value of the *i*-th user to the *j*-th object):

$$r_{ij} = \frac{k * a_n + (k+1) * a_{n-1} + \ldots + a_{n-k}}{\Sigma_{i=1}^{k} i} \tag{5}$$

The values given to the object at different times and in different sources by the user is taken into consideration. This formula prefers the latest value given by the user.

Analyzing the passport data of objects can also provide valuable information. Included information is the warranty period and duration for objects, things, and services. First of all, frequency of occurrence of the name of the object as the results of trade operations, advertising sites, comments by users, print products and in other digital and non-digital advertising methods may be included here.

If the name of the analyzed object in the different methods is greater than others, then this frequency may indicate the importance of the object for users. But sometimes one of the two objects frequencies can be much higher than the frequency of the other. Then, it may be important to reduce the difference between them.
For this purpose, it is possible to use the following logarithmic transformation.

$$\omega_{t,d} = \begin{cases} 1 + \log_{10} tf_{t,d}, tf_{t,d} > 0 \\ 0, \textit{in the other cases} \end{cases}$$

So, we can formulate a matrix from the value given by the users to the objects (things, products, services, etc.) for the solution of the problem by making some simplifications:

$$R = (r_{i,a})_{i=1,a=1}^{MN} \tag{6}$$

Here, $r_{i,a}$- is the value given by the i^{th} user to object "a". In this matrix the number of users is M, the number of objects is N. In each row, the values given to different objects by the user are shown, and the value given to an object by different users is shown in each column. In a specific line of this matrix, for example, generated data from the passport base can be placed in the last line. In this case, final values given by producer firms to their products can be in the row M + 1. As a result of the system's work, adequacy rate of producer firms to market can be calculated. After the value line of producer is added, there will be column N and row M + 1 in the matrix R. It can be considered for simplicity that there is still a line M on the matrix R. In this matrix, each column and each row are vectors. Similar operations can be applied on these vectors. The proximity of two vectors can be calculated by various methods.

Euclidean distance, which is a proximity measure of two vectors, cosine metric, and correlation coefficient of Pearson, Tanimoto coefficient, and Manhattan distance and so on, can be used. In most cases, cosine metric Tanimoto correlation measures and Pearson correlation coefficient are used in the recommendation systems.

In the study, mainly Pearson correlation coefficient and cosine proximity measures are used. The obtained results are used in the elements of matrix R. Pearson correlation coefficient, which is the proximity measure between two users, is calculated in two vector rows (vector rows i, j) by this way:

$$sim(i,j) = \frac{\sum_{p\in P}(r_{i,p} - r_{i,a})(r_{j,p} - r_{j,a})}{\sqrt{\sum_{p\in P}(r_{i,p} - r_{i,a})^2} * \sqrt{\sum_{p\in j}(r_{j,p} - r_{j,a})^2}} \qquad (7)$$

Here i, j – show users, p – set of product or services

$r_{i,p}$ – shows value given by user i to product, goods or service p

$\boldsymbol{r_{i,a}}$ – shows the average value given by user i to objects, products or services $\boldsymbol{p} \in \boldsymbol{P}$

If there are objects in $\boldsymbol{N_p}$ number of in P set, it can be calculated as follows:

$$\boldsymbol{r_{i,a}} = \frac{\sum_{p\in P} r_{i,p}}{N_p} \ .$$

$\boldsymbol{r_{j,p}}$ –shows value given by user j to objects, products, or services p

$\boldsymbol{r_{j,a}}$ –shows the average value given by user j to objects, products or services $\boldsymbol{p} \in \boldsymbol{P}$

This average value can be calculated as follows:

$$\boldsymbol{r_{j,a}} = \frac{\sum_{p\in P} r_{j,p}}{N_p}$$

The proximity of the values given by different users to two objects (product, services) can be given as cosine proximity measure between two columns of vectors.

$$\omega_{a,b} = \frac{\sum_i r_{i,a} r_{i,b}}{\sqrt{\sum_i r_{i,a}^2}\sqrt{\sum_i r_{i,b}^2}} \qquad (8)$$

Measure of matrix (6) should be reduced to achieve successful result of the recommendation system. For this purpose, we can select appropriate rows which are suitable users close to certain user, and appropriate columns that are close to the object of interest to the user. Therefore, if the interest of the ith user to object c is known, appropriate rows can be recommended to the user who is close to the ith user and objects in the intersection of columns close to object c and it can be presented about selected products. As in many practical applications, proximity measure between the rows can be taken as Pearson proximity measure and cosine proximity measure. The solution of the problem can be given as rows, which is following the given conditions (users close to the ith user - j), and columns (objects close to services or products - b).

$$sim(i,j) = \frac{\sum_{p\in P}(r_{i,p} - r_{i,a})(r_{j,p} - r_{j,a})}{\sqrt{\sum_{p\in P}(r_{i,p} - r_{i,a})^2} * \sqrt{\sum_{p\in j}(r_{j,p} - r_{j,a})^2}} \le \Delta sp \qquad j \in [1, M] \qquad (9)$$

$$\omega_{a,b} = \frac{\sum_i r_{i,a} r_{i,b}}{\sqrt{\sum_i r_{i,a}^2}\sqrt{\sum_i r_{i,b}^2}} \le \Delta\omega \qquad b \in [1, N] \qquad (10)$$

Here:

Δsp – is the parameter that defines the width of selected rows. There may be the i[th] row that is suitable for these conditions in very small values of this parameter.

$\Delta\omega$ - is the parameter that defines the width of selected columns. This parameter can also provide a column for solving the inequality in smaller values.

We are interested in users close to the i-th user and the objects close to the object *c*, and assume that the system solutions are users and objects that make up the following pairs

$$\{(i_1, a_1), \ldots, (i_l, a_l)\} \qquad (11)$$

From the given expression (11) it is known that, at least one user and one object is included into the set of selected users and objects. So, solution of the problem is $\{i_1, i_2, \ldots, i_l\}$ - a group of users and $\{a_1, a_2, \ldots, a_l\}$ a group of objects.

5. CONCLUSION

This article highlighted the development stages of creation, the purpose, and principles of recommendation systems. At different stages of the development of the recommendation systems, it was determined software tools which serve to personalized interest and corporate interests of large companies.

In this article, several approaches to the creation of recommendation systems were studied, method and algorithm for hybrid recommendation systems based on obtained through different methods were given. It is possible to determine the semantic conformity between the user and user group who want to obtain the product in a virtual space. Also, it is possible to define the similarity between the product that the user is interested in and the other products that fit the parameters. The results of this study can be used to evaluate the product, determine a group of users who want to get specific products, and other practical issues related to marketing.

Funding: This study received no specific financial support.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.

REFERENCES

[1] S. Rahul, c. Kanika, and R. Akshama, "A survey on the generation of recommender systems," *International Journal of Information Engineering and Electronic Business*, vol. 9, pp. 26-35, 2017.Available at: 10.5815/ijieeb.2017.03.04.

[2] F. Ricci, "Lior Rokach. Bracha Shapira. Paul B. Kantor. Recommender system handbook," ed. New York Dordrecht Heidelberg London: Springer, 2011.

[3] K. K. Hashimova, "Analysis method of internet advertising-marketing information's dynamic changes," *International Journal of Information Engineering and Electronic Business*, vol. 9, pp. 28-33, 2017.Available at: https://doi.org/10.5815/ijieeb.2017.05.04.

[4] D. L. Richard, *Theory of organization*. M: Unity-Dana, 2006.

[5] Kartiev S.B. and O. I. Sharshakov, "Development of recommendation systems using the model of spark calculations," presented at the All-Russian Scientific and Technical Conference Functional and Applied Aspects of Computer Technologies and Information Security. Materials of the Conference, Taganrog, 03-09 April 2017.

[6] E. A. Breykin, "A recommended system based on collaborative filtration," *Young Scientist*, vol. 13, pp. 31-33, 2015.

[7] G. Costa and R. Ortale, "Model-based collaborative personalized recommendation on signed social rating networks," *ACM Transactions Internet Technology*, vol. 16, pp. 1-21, 2016.Available at: https://doi.org/10.1145/2934681.

[8] V. Vellaichamy and V. Kalimuthu, "Hybrid collaborative movie recommender system using clustering and bat optimization," *International Journal of Intelligent Engineering and Systems*, vol. 10, pp. 38-47, 2017.Available at: https://doi.org/10.22266/ijies2017.1031.05.

[9] D. E. Koroleva and M. V. Filippov, "Analysis of learning algorithms for collaborative advisory systems," *Engineering Journal: Science and Innovation*, vol. 6, p. 6, 2013.

[10] K. K. Gashimova, "Determining the quality of the site based on customer reviews," *International Education Institute Educatio*, vol. 23, pp. 40-43, 2016.

[11] E. D. Michael, R. T. John, and K. A. Joseph, "Collaborative filtering recommender systems," *Foundations and Trends in Human–Computer Interaction*, vol. 4, pp. 81–173, 2010.