2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science

User interest prediction model based on the Cellular Automata

Yecheng Wang
School of Science, Jiangnan University
Wuxi, Jiangsu, China
wyc_993@163.com

Xu-Qing Tang
School of Science, Jiangnan University
Wuxi, Jiangsu, China
txq5139@jiangnan.edu.cn

***Abstract*—Utilizing the concept of the Cellular Automata, this paper builds the user interest model. In this model, the user is considered as a cell in two-dimensional lattice, the degree of interest is used as the state of a cell, and the factors affecting the user's interest are quantified, which is used as the cell's movement rules. This model can be used to simulate the change of interest in something after the user is influenced by others and by time. Simultaneously, the model is improved and optimized by taking into account the impact of emergencies and the sudden increase in the number of individual users on user interest. Finally, using the real data of the iPhone mobile phone sales, this paper verified the rationality and effectiveness of the model.**

Keywords- user interest; cellular automata; prediction model

I. INTRODUCTION

With the development of network technology and the improvement of people's ideological level, network information has entered an era of explosive development, and a large amount of information has flooded into the network, making the amount of network information present a geometric progression, including many non-structures information and duplicate information [1]. In this context, the network information presents a huge but scattered, and disorderly character, which greatly increases the difficulty for people to search and use information effectively. At the same time, the level of network users is uneven, especially for those who are unfamiliar with search methods, and who do not have deep retrieval capabilities. It is difficult to find the required information [2]. In order to solve the above problems, search engines and recommendation systems have gradually developed with the intensification of network information. Google, Baidu and other search engines facilitate people's lives and meet people's spiritual and cultural needs.

People enter the information that needs to be searched in the search engine, and then enter the search result recommendation page, afterwards, the user selects the information he needs in the numerous results. Some Web pages have also developed a scoring system. After the user completes the search process, he or she will score whether the search results meet the requirements. During the entire search process, the cookie stores the user's relevant information in the database, such as browsing logs, user registration information, and historical scoring records, in order to increase the speed of the user's next search [3]. The recommendation system analyzes these information and discovers the interest bias in the user's search process, so that in the next search result recommendation, the user's search interest can be combined to implement personal interest information recommendation. If the user's interest needs change, the recommendation system will adjust the recommended content in time to achieve a truly personalized recommendation service. The recommendation system currently has a very wide range of application prospects. Taobao, Netease Cloud, Amazon, eBay, and many social networking sites all use various forms of recommendation systems in various degrees and have achieved greater benefits. For example, according to Venture Beat statistics, Amazon's recommendation system provides 35% of merchandise sale [4].

The general recommendation system model flow is shown in Fig. 1.

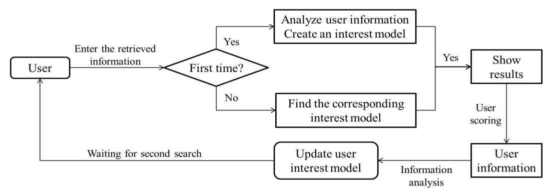


Figure 1. General recommendation system model flow chart

As can be seen from Fig. 1, the user's interest model plays a crucial role in the overall recommendation system. Only by establishing an accurate user interest model can corresponding recommendations be made to according to the model, thereby improving the accuracy of the recommendation.

Fu et al.[5] used internal drive theory to propose a linear regression model to describe the relationship between Web users' browsing behavior and interest degree from the perspective of psychology; Xing et al. [6] focused on the particularity of users' reading time and proposed the method of evaluation interest values based on mixed behaviors; Liu et al. [7] used RBF neural network model to quantify user interest degree; Xia et al. [8] introduced interesting degree into vector space model and expressed it in two-level tree structure. In this paper, the user interest model is established by using the method of cellular automata through the study of changes in user interest. This model starts from the user's psychology of herd and considers the firmness of the user's own will, so as to explore the development trend of the public interest of the user and help improve the accuracy of the recommendation system.

II. BASIC COMPOSITION OF USER INTEREST CELLULAR AUTOMATA

Cellular automata (CA) [9] is originated from a von Neumann thought experiment on machine reproduction, which led to CA's pioneering work. Later in the 1970s, J. Conway compiled a game program called "Life", and people's passion for studying cellular automata was once again inspired, and shown that the cellular automata is

978-1-5386-7445-1/18/$31.00 ©2018 IEEE
DOI 10.1109/DCABES.2018.00057

equivalent to the Turing machine, that is, when given the appropriate initial conditions, the cellular automata can simulate any kind of computer. At present, cellular automata are widely used in various fields of social, economic, military, and scientific research.

CA is a dynamic system that is defined in a cellular space composed of discrete, finite state cells and evolves in discrete time dimensions according to certain local rules. Its most basic components include Cell, Lattice, Neighbor, and Rule.

A. Cell and its state space

The definition of the user interest model is established in the two-dimensional cell space, and each cell represents one user. Users have two attitudes for searching an item: interest and lack of interest. This article uses the value range of interest to indicate the user's attitude towards the project.

For a certain user i, $ID(t_i)$ is used to represent the interest degree of cell *i* at time *t* at a certain item.

$$ID(t_i) = \frac{project - score}{scoring - limit} \tag{1}$$

The users' scoring of the project is obtained through the display of acquired information or through the use of data mining technology from the user's actual browsing behavior, such as user inquiries, browse time on the page, favorites, and copy operations when accessing the page. These information can be quantified as the user's score on the project [10].

From (1), it can be seen that $ID(t_i) \in [0,1]$. The bigger $ID(t_i)$ is, the higher the user *i*'s evaluation of the project is, which means it has a higher interest in the project. Each cell user has his or her own interest in something, but only when the interest reaches a certain threshold, he or she will actually pay for interest. In real life, the threshold of interest is affected by factors such as the level of economic development. For example, when a user's interest in a book is 0.6, he or she may choose to purchase. When the user's interest in a luxury product is 0.6, there is a great probability that it will not be purchased. Mark interest threshold as β, and makes the following provisions:

When $0 \le ID(t_i) < \beta$, cell i is not interested in this project. The smaller the $ID(t_i)$, the lower the interest;

When $\beta \le ID(t_i) \le 1$, cell i is interested in this project. The larger the $ID(t_i)$, the greater the interest.

B. Neighbor

This paper defines the neighborhood of the cell as eight neighbors. Because cyberspace is a virtual world, neighbors in cyberspace are not adjacent to "spatial distance" but are adjacent to "psychological distance". Therefore, this paper uses distance between individuals to determine distance [11].

The change in the interest of a public user in a certain thing is often attenuated or increased according to a certain rule, so it can be considered as a periodicity. Therefore, the boundary condition defining cyberspace is periodic. The periodic boundary conditions refer to the relative connections to each other. For two-dimensional space, they are connected up and down, and left and right [12].

C. User interest changes rules

The psychologist K. Lewin used the force field theory [13] to point out that human behavior is influenced by its own subject and environmental object, and proposes the basic formula of human behavior: $B{=}F(PE)$. Where *B* is human behavior, *P* is the main body, and *E* is the object. The psychologist Hilgard put forward the theory of internal drive based on the research of Hill [13]. Fu et al. [5] applied internal drive to analyze the user's behavior, and believed that the user's current decision-making depends on the result of the user's past acceptance of the stimulus. Therefore, for a user of a cell, the change in the degree of interest is mainly influenced by "Neighbor" and time, as well as some other minor influences. Below, this paper will start with these two major influencing factors and establish rules for changing user interests.

1) Influence from neighbors

The notation that defines the central cell is the cell (i, j).

Assume that the cell (i, j) has a certain interest in a certain item at time t, and marked as $ID(t_{ij})$. Then $ID(t_{ij})$ is affected by the surrounding eight neighbors. The interest degree change of the cell (i, j) at time *t* is defined as follows:

$$ID(t_{ij})' = \sum_{\substack{i-1 \le m \le i+1 \\ j-1 \le n \le j+1 \\ (m,n) \ne (i,j)}} ID(t_{ij}) \times \frac{ID(t_{mn}) - ID(t_{ij})}{ID(t_{mn})} \tag{2}$$

From (2), it can be seen hat when the interest degree of neighbor cells is greater than that of the central cell, $ID(t_{mn}) - ID(t_{ij}) > 0$, the attitude of the neighbor will have the same effect on the central cell, on the contrary, it will have a reverse effect.

2) Influence from time

The interest degree of a cell for a project decrease with time, and it can be assumed that the relationship between the decay of interest and the time satisfies the exponential function and is defined as follows:

$$ID((t+1)_{ij}) = (1 - 1/\lambda)^{ID(t_{ij})'} \tag{3}$$

where λ is the firm coefficient of the central cell (i, j) and $\lambda > 1$. It can be seen that the larger the value of λ, the slower the trend of the interest degree of the cell (i, j) over time, which is in line with person's character reality. When the person's will to a certain thing is firmer, the slower his interest in that thing will change.

III. User Interest Prediction Model and Optimization

A. User interest prediction model

Changes in the user's interests will be affected by neighbors and will change over time. In the process of changing with time, once the state of user interest at time *t* is known, the state at time *t*+1 can be obtained without resorting to the user interest state before time *t*. According to this characteristic, the change process can be considered as Markov process, and because the user's interest state and time are discrete, this process can also be called the Markov chain [14]. Combining the user's interest with the neighbor (Equation (2)) and time (Equation (3)), a prediction model of user interest can be obtained:

$$\begin{cases} ID(t_{ij})' = \sum\limits_{\substack{i-1 \le m \le i+1 \\ j-1 \le n \le j+1 \\ (m,n) \ne (i,j)}} ID(t_{ij}) \times \dfrac{ID(t_{mn}) - ID(t_{ij})}{ID(t_{mn})} \\ ID((t+1)_{ij}) = (1 - 1/\lambda)^{ID(t_{ij})'} \end{cases} \tag{4}$$

In (4), the interest degree $ID((t+1)_{ij})$ at time $t+1$ can be obtained from the initial interest degree $ID(t_{ij})$ of the cell (i, j) at time *t*, and then through the loop iteration, the user's interest $ID((t+k)_{ij})$ for time $t+k$ can be obtained, $k = 1, 2, \cdots, n$.

B. User interest prediction model optimization

As mentioned earlier, for a cell user, the change in interest is mainly influenced by neighbors and time, but it may also be influenced by some other secondary factors. In the following, this article will be based on the prediction model, consider the impact of unexpected events and the increase in the number of users on the user's interest, and further optimize the model.

1) The impact of unexpected events on user interest

When there are some incidents in the community, some people's interest in certain projects will change. For example, Lotte suffered 500 billion won (approximately RMB 3 billion) in damage due to the Sadr incident. However, this effect will often only last for one or two periods and will not last forever. This effect occurs on the time axis, therefore, add an influence factor $\alpha(\alpha > 1)$ to (3), and then get a new time decay interest degree:

$$ID((t+1)_{ij}) = \alpha \times (1 - 1/\lambda)^{ID(t_{ij})'} \tag{5}$$

From (5), it can be seen that when α increases, the user's interest increases.

2) Increase in the number of individual users

Popular users generally have herd mentality. When many people are pursuing a certain thing, the thing will have a certain topic degree. The higher the topic degree, the more it will attract new users to participate. That is, the more attractive the project is, the more users will be interested. Use $AD(t)$ to represent the attraction degree of an item at time t to the user.

$$AD(t) = N(t) / N_{all} \tag{6}$$

where $N(t)$ indicates the total number of users whose interest degree is greater than the interest degree threshold, and N_{all} indicates the total amount of users in the network. The new user is considered that $ID(t_{ij}) \in [\beta, 1]$, that is, the user will search only if they are interested in the project.

IV. Results and Analysis

In order to verify the rationality and effectiveness of the model, this article selects the sales of the iPhone smartphones that have been popular since 2007 for simulation experiments.

In 2007, the first generation of iPhones is introduced by Apple. Once it was on sale, it triggered an upsurge of resentment. With the enhancement of the capabilities of the new model, as well as the improvement of people's living standards and purchasing power, the sales volume of iPhone handsets has also been increasing, especially when the new generation is launched, sales volume will have a sudden increase.

On the Statista website, iPhone sales for each quarter between the third quarter of 2007 and the first quarter of 2018 can be found. (Data sources: https://www.statista.com/statistics/263401/global-apple-iphone-sales-since-3rd-quarter-2007/)

Combining with the original data, it can be seen that the sales volume of iPhone handsets has shown a very obvious upward trend, and in the first quarter of 2012-2018, there will be a sudden increase in mobile phone sales. After reviewing the information, since the beginning of 2012, iPhone's release time is basically between September and October, according to the US fiscal year, that is the first quarter of each year.

According to the known data, the sales volume is roughly distributed between 0-79 (million). Suppose that each sales volume represents a user whose interest degree exceeds the threshold. Then in Cellular Automata, a grid is used to represent a cell user. In order to ensure that the highest sales volume can be reflected in the network space and due to the limitation of the memory capacity of personal PC, the size of the network space is set to 89×89.

In this simulation experiment, user interest data is difficult to obtain, but according to the basic trend of human social development, under a large number of users, it can be considered that users' interest in something is in a 0-1 normal distribution. Therefore, the initial value of interest of all users is randomly given in accordance with the normal distribution of 0-1.

In the iterative process, taking into account the increase in people's purchasing power, it is believed that the user interest threshold increases slightly with the number of iterations.

The release of the new mobile phone will inevitably trigger a wave of purchases, so when the number of iterations is iterated to several points at *T*=19,23,27,31,35, 39,44, which is in the first quarter of 2012-2018, add the influence factor and restore the initial value after iterating once.

However, since the initial value of the user's degree of interest is given at random, which neglects the development of the users' personality, in order to make the prediction model as accurate as possible, the concept of error value σ is introduced, and the value σ is calculated as:

$$\sigma = \sqrt{\frac{\sum_{k=0}^{T-1} [N_p(t+k) - N(t+k)]^2}{n}} \tag{7}$$

where $N_p(t+k)$ is the sales forecast value of the *k*-th period, $N(t+k)$ is the actual sales data for the *k*-th period, and *T* is the total number of forecast periods. In this example, *T*=43.

After eight repeated tests, eight σ values were obtained as shown in TABLE I.

TABLE I. The value of σ after repeated tests

The value of σ	
6.5894	6.4635
6.4793	6.4117
6.4190	6.1904
6.1199	6.4255

From the results in TABLE I, it can be seen that the value of σ is relatively small and maintained at a very stable level, so the model can be considered to have strong stability. Selected the result with the smallest σ value for simulation, the result of the comparison between the simulation data and the original data was shown in Fig. 2.

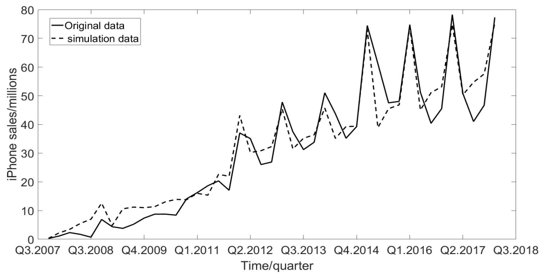
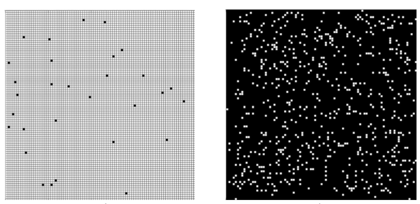


Figure 2. User interest model simulation results

It can be seen from Fig. 2 that the sales volume has basically risen in the 11 years and there has a sudden increase in the quarterly sales volume with an increase in the impact factor, and then dropped to normal levels. The simulation results of this model are basically consistent with the real statistics. The corresponding cellular automata iteration results are shown in Fig. 3.



Iteration 1 Iteration 30

Figure 3. Cellular Automata Iteration Results

In order to more clearly show that the user interest degree is affected by both time and neighbors, this paper separately removes the single influencing factors and makes a fitting graph. The simulation results are shown in Fig. 4.

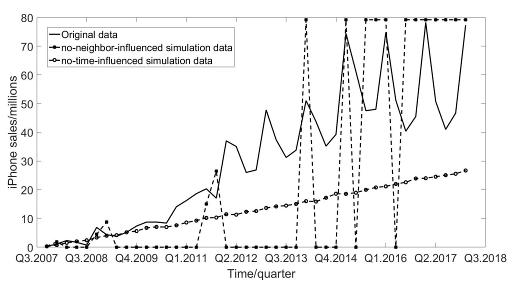


Figure 4. User interest model simulation results after eliminating neighbor influences and eliminating time effects

From Fig. 4, it can be seen that after eliminating the influence of neighbors and time respectively, the sales trend of the simulation and the original data have a greater difference, once again proved that the model is indeed mainly affected by the dual impact of neighbors and time.

V. CONCLUSIONS

Cellular Automata is essentially state transfer functions for the cell in time and space. The change in user's interest in something is essentially a change of state over time and other influencing factors. Therefore, it is universally applicable to use Cellular Automata to simulate changes in user interests. This paper starts from the two main factors affecting the user's interest and establishes a model of user interest changes. Using this model, the trend of change in the interest degree of users can be understood in advance, so that the appropriate measures can be taken to guide and achieve mutual benefits.

However, in the process of establishing the model, the determination of the model parameters does not set a uniform standard, and can only be adjusted through the existing data. How to carry out the unification of parameter standards still needs further study.

VI. ACKNOWLEDGE

This work is supported by the Undergraduate Innovation Training Program of Jiangnan University of China Grant NO. 201710295077 and National Natural Science Foundation of China Grant No.11371174.

REFERENCES

[1] R. Huang. Network Information Retrieval and Utilization. Wuhan: Wuhan University Press, 2002.4

[2] J. Chen. Problems in Network Information Retrieval and Countermeasures. Library, 2007(05):125-126.

[3] Y. Ma, H. Qian, Y. Sun. Research on Application of Cookie in Web Authentication. Miniature Microcomputer Systems, 2004(02):207-210.

[4] J. Liu, T. Zhou, B. Wang. Research progress of personalized recommendation system. Advances in Natural Science, 2009, 19(01):1-15.

[5] G. Fu, Z. Zhu. User Interest Modeling Based on Behavior Analysis in Personalized Service. Computer Engineering & Science, 2005(12):80-82.

[6] L. Xing, Z. Song, Q. Ma. User interest model based on mixed behavioral interest degree. Journal of Computer Applications, 2016, 33(03): 661-664+668.

[7] J. Liu, P. Sun, H. Ni. Estimation of User Interest Based on Neural Network. Computer Engineering, 2011, 37(07): 187-189+192.

[8] Y. Xia, Y. Liu. An improved method of user interest calculation and user interest correction. Modern Information, 2014, 34(01):46-48+55.

[9] W. Fang, L. He, K. Sun, et al. Research on the model of network public opinion propagation using cellular automata. Journal of Computer Applications, 2010,30(03): 751-755.

[10] J. Li, W. Guo, Z. Tang. User interest calculation based on user browsing behavior analysis. Computer Engineering and Design, 2012, 33(03): 968-972.

[11] I. Ray, T. Xin. Implementing Real-Time Update of Access Control Policies[C]//Research Directions in Data and Applications Security Xviii, Ifip Tc11/wg 11.3 Eighteenth Conference on Data and Applications Security, July 25-28, 2004, Sitges, Catalonia, Spain. DBLP, 2004:65-80.

[12] X. Zeng, Y. Fang, Y. Yuan, et al. Network public opinion incentive model based on cellular automata. Computer Applications, 2007(11):2686-2688+2714.

[13] C. Hu, H. Qiao. Information Service and Users. Wuhan: Wuhan University Press, 2001.

[14] X. Deng, Y. Deng, Y. Zhang, et al. A Reliability Markov Model and Its Application.Acta Automatica Sinica, 2012, 38(04):666-672.