2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)

Internet Advertising Investment Analysis Based on Beijing and Jinhua signaling data

1st Xia Song
Wuhan University
Electronic information school
Wuhan, China
1024316809@qq.com

2nd Jiang Hao
Wuhan University
Electronic information school
Wuhan, China
jh@whu.edu.cn

3rd Zhang Yi
Wuhan University
Electronic information school
Wuhan, China
zyi@whu.edu.cn

4th Peng Duo
Wuhan University
Electronic information school
Wuhan, China
1595106631@qq.com

***Abstract*-With the popularization of the internet, the media of advertising tends to diversify. Compared to traditional advertising methods, internet advertising is well received for its convenient and fast propagation mode. However due to its mandatory and invasive property, the publicity effect of internet advertising is greatly reduced. To make the delivery of Internet advertising more directional and efficient, we try to analyze the behavioral preferences of netizens from Beijing and Jinhua when using different Apps, and then conclude some rational delivery strategies for internet advertising investors. We use Multi-state Models (MSM) to construct an online Apps transfer model for netizens, which will help to get some intuitive conclusions. After that, basing on online apps transfer model, we adopt entropy method, hierarchical clustering, and Tucker decomposition to mine the potential behavioral preferences of netizens. Finally we provide some suggestions and references with the internet advertisers.**

Keywords-Internet Advertising, signaling data, surfing preference, data analysis

I. INTRODUCTION

It is hard to avoid mention of Big Data anywhere we turn today [1]. There is a broad recognition of the value of data, and products obtained through analyzing it [2]. Popular news media now appreciates the value of Big Data as evidenced by coverage in the Economist [3]. According to the research, the global IP traffic will triple in 2016-2021. Global Internet users will grow from 3.3 billion in 2016 to 4.6 billion, accounting for 78% of the global population. Under this situation, in china the number of mobile phone users has reached 788 million, and the proportion of mobile phone users was as high as 98.3%. Obviously, the internet, an indispensable part of modern life, is constantly infiltrating into all aspects of human work and life.

The Internet can provide users with large and valuable information, thus it also provides an efficient way for advertising. Advertising, an important way to show the value of goods and services to customers, can stimulate consumers' desire to buy [4] [5]. Thus Companies are pouring billions of dollars into Internet advertising to obtain greater return on investment on ads [6], which contributes to the appearance of Internet advertising industry. Internet advertising is a high-tech advertising operation method where investors advertise on the Internet through online advertising platforms, use advertisement banners, text links, and multimedia methods on the website to publish advertisements on the Internet and deliver them to Internet users through the network.

But nowadays the internet advertising industry is not mature enough. We examine three latent variables of Internet ad avoidance: perceived goal impediment, perceived ad clutter, and prior negative experience. Perceived goal impediment is found to be the most significant antecedent explaining advertising avoidance on the Internet [7]. How to make internet advertising more targeted is a problem demanding prompt solution. Accurate marketing which sends the right information to the right customers through the right channels at the right time can improve the accuracy of delivery, thus truly affecting the purchase decision of the target customers and promoting the purchase behavior. The effectiveness and accuracy of internet advertising can be improved significantly by integrating the accurate marketing into the procession of internet advertising.

II. RELATED RESEARCH

At present, many scholars have carried out researches on systematically focusing on the effectiveness of mobile internet advertising marketing. American media theorist Paul Levinson explores the pros and cons and development prospects of mobile media from a multi-dimensional media perspective. The author believes that mobile media will lead to a revolution in information transmission, and mobile media will dominate the way of information dissemination in the future [8]. Under this circumstance, Meikang Qiu proposes a way which can both minimize the energy cost and satisfy the efficiency need when processing the multimedia information [9][10].American scholar Cindy Krum made a panoramic description of the development of Internet marketing in the United States in his 2012 book [11], providing many scholars with an overall impression of mobile Internet advertising marketing and the path of related research. In the book "The third screen: marketing to your customers in a world gone mobile", Chuck Martin conducted a detailed study on the terminal forms, marketing methods and applicable industries of mobile internet marketing, indicating an effective way for companies and customers to deepen their understanding through the Internet and providing a theoretical basis for the development of Internet advertising [12]. Chinese scholar Yu Kun analyzed the

978-1-7281-1664-8/19/$31.00 ©2019 IEEE
DOI 10.1109/CSE/EUC.2019.00084

factors influencing users' acceptance of Internet advertising based on the UTAUT model [13], and found that the consumer experience has a huge impact on the acceptance of advertising. Overall, the above research focuses on the theoretical construction of Internet marketing benefits, and demonstrates the enormous potential and commercial value of Internet marketing from different aspects. In the Internet marketing model, the user's behavioral orientation plays a key role in the effectiveness of advertising, but this model did not specify a way to explore the behavioral orientation of users under the influence of the Internet, lacking comprehensive guidance on the benefit analysis of the specific investment mode in practice.

Many scholars have drawn relevant conclusions on the study of people's behavioral orientation under the influence of the Internet. In communication studies, Melowitz believes that the emergence and application of new electronic media has brought about the fusion and interaction of different scenes. And different scenes have finally changed the behavior pattern of the public [14]. So the study of different scenes has profound implications for mining the user's behavioral orientation. Eric J. Johnson argues in Digitizing Consumer Research that the factors which dominate consumption in the information age are not just the original cognitive and theoretical systems of time, input costs, and product value creation [15]. The rapid spread of information will has a potential impact on the way people behave, and in this situation the purchasing and consumption tendencies are also unconsciously changing. Sun Haiyang's "Design of scenes interaction in mobile consumption" proposes that the development of China's mobile consumer sector has shifted from functional consumption to scenes consumption, which means that the competition of products change from the competition to meet the user's utility needs to the consumption scenario solution. The construction of consumption scenarios is the core strategy for the development of the mobile consumer industry now and in the future [16]. The above scholars have analyzed the online behavioral orientation of the masses in different scenarios, but most of them focus on theoretical research and most of the behavioral orientations of different scenarios in the real population are collected by questionnaires and interviews. The size, the subjective factors, and the representativeness of the population will have a certain negative impact on the conclusion.

For the above research, the survey sample set is not large enough. The actual guidance method is vague, and the statistical survey results may be subjective. The network signaling data coming from the objective and full sample is used in the paper to analyze the characteristics of the crowd, and we combine the actual needs to obtain the operational strategy of Internet advertising, which ensures the objectivity of the results. This kind of big data analyzing ways are widely used in nowadays society. For instance, in the Medical field, Jihe Wang enable real-time information service on telephone system [17]; Min Chen propose the Smart Personal Health Advisor [18]. Both of them are based on big data platform and get great success. Thus, to get more thorough conclusion, we adopt the similar way, and in the course of signaling data processing, we use MSM to filter and integrate data, to turn the disordered raw data into order and useful data. The MSM used to integrate the individual state transition data from the signaling data is sensitive to the state transformation. After that we take the individual state transition data as the basic data for studying the behavior transfer mode of netizens. To judge the richness and research value of the selected data, we use entropy method, a mathematical analysis method to evaluate the dispersion of the data. Meantime, in the behavior analysis process, the entropy value can also map the transformation characteristics to the population behavior pattern. Therefore, the above two methods are used to complete the pretreatment of the data. In order to do further analysis on the online behavior patterns of the two groups of people, clustering algorithm, a method which is useful to segment several groups with similar behavior from the sample database, can be used to divide the crowd into several feature classes with similar behavior patterns, which is helpful for intuitively exploring the online behavior characteristics of the crowd. The clustering problem has been addressed in many contexts and by researchers in many disciplines, which reflects its broad appeal and usefulness as one of the steps in exploratory data analysis [19]. Hierarchical clustering, a kind of clustering algorithm which is not sensitive to the input order of samples, can obtain multi-level clustering structure at different granularities. It has a big advantage in dividing behavior transfer mode. Tian Tian, Jun Zhu, through hierarchical clustering, accurately identify people with similar online style characteristics from Internet data [20]. The Silhouette Coefficient combining the two factors of cohesion and resolution can be used to judge the quality of the clustering results. It is proposed by Peter J. Rousseeuw in 1986[21]. In this experiment, the sample is divided into several classes with similar behavior by hierarchical clustering, and the Silhouette Coefficient is used to judge whether the result is satisfying. Subsequently we outline the habits or behavior patterns of each type of class. In order to eliminate the uncertainty of the clustering result algorithm itself, we use Tucker Decomposition to verify the clustering result. Tucker Decomposition reduces the diversity of behavior patterns of the crowd, which is convenient for intuitive analysis and has a good effect in the study of population behavior patterns. When studying the interpersonal relationship, Aaron Schein and Mingyuan Zhou used Tucker's decomposition to analyze the interaction data between the communities and then explore the intuitive and potential characteristics of interpersonal communication [22]. In this experiment, the above methods are used to deeply explore the behavioral orientation of different consumption levels in different scenes modes. By analyzing the behavioral orientations of different Internet scenes modes, we can integrated the rational delivery strategy of Internet advertising of different types of products.

III. DATA SOURCE AND SYSTEM MODEL

A. Data source

We choose the netizen in Beijing and Jinhua as the sample. As the capital of China, Beijing is the political and cultural center of the country, whose population is dominated by young and middle-aged consumers. As a city with rapid economic development and technological innovation, Beijing's consumer market is very broad. As a prefecture-level urban area in Zhejiang Province, Jinhua is a city with relatively low level of commercial development. Therefore, the behavior orientation of consumers in the two cities is studied, which makes our research more representative.

The main data packets used in this study are derived from mobile phone signaling data from Beijing and Jinhua City. The selection period is 23 days and the number of users covers about 4 million. The data mainly includes information such as user id (encrypted), user equipment, session start and end time, and url.

B. System model

In order to reveal the behavioral orientation of different user groups when using various types of apps, we adopt the MSM model, Entropy Calculation, Hierarchical Clustering, Tucker Decomposition and other analytical methods. Firstly we take personal transfer relationship of the apps as the basis data. After that we represent the characteristics of online behavior of groups by the characteristics of classes. The characteristics of online behavior of groups is then mapped to a specific kind of people with similar behavior. Finally we plan the targeted advertising strategy for Internet advertising by analyze the online characteristics of all kinds of people.

1)MSM model preprocesses data

The MSM is commonly used to analyze continuous-time Markov models, which is usually determined by the state of the moment and the state strength of the next moment. The state can be changed at any time during processing. The likelihood of a simple continuous-time Markov model is usually determined by the transfer strength matrix Q, which usually represents the initial value between states allowed to transfer in the Markov chain [23]. When processing the original signaling data, the transfer between the internet states is not sorted according to the individual. Therefore, in the MSM model, the individual is regarded as the main analysis object, and the different states of the individual are sorted according to the time sequence. In this situation, different categories of apps mean different states of individual. So in the course of processing, we count the order and frequency of the individual's usage of apps at different nodes, and then we adopt MSM and by transfer intensity Q matrix we integrate the transfer relationship of individual apps usage in continuous time states. Finally we get the state transfer model of the crowd when using apps, and the pretreatment of the data is finished.

2)Entropy Calculation evaluates the value

Entropy Calculation, an efficient method for generating high-resolution density distribution of data from a finite number of diffraction data [24], is useful to evaluate the value of the data. After deriving the users states transfer model, we adopt the Entropy Calculation to objectively evaluate the diversity of the data. Conditional Entropy is one of the objects of Entropy Calculation, and the expression is $H(x|y)$, which means the uncertainty of event x after knowing event y. Since the state sequence has different lengths between experimental samples. The single Entropy Calculation will ignore this state transition characteristic. Conditional Entropy can capture this kind of property, and its calculation formula is (different states is treated as a queue)

$$H(x|y)=E[\log\frac{1}{p(a_j|b_i)}]=-\sum_{j=i}^{q}\sum_{i=1}^{q}p(a_j|b_i)\log p(a_j|b_i) \qquad (1)$$

In the formula, q indicates the length of the queue, and i , j is the serial number of the current state. b_i indicates that the current state of the individual is the i-th online status in the queue, and a_j indicates the state of the next moment is in the j-th online status in the queue . $\log p(a_j|b_i)$ indicates the probability of transitioning from b_i state to a_j state. The smaller the probability is, the larger the conditional entropy value is, which indicates the sample is diverse. So we calculate the conditional entropy of the transfer model of Beijing and Jinhua. In order to make the result more intuitive, we normalize the result.

3)Hierarchical Clustering divides the class

On the basis of the state transition model, the Hierarchical Clustering method is used to aggregate people with similar online styles from the online state transition model. Hierarchical Clustering can integrate data with similar characteristics from the dataset, usually used to explore some inherent structures of the data itself [25]. The following are the steps of hierarchical clustering. At the beginning, we regard individuals in state transition models as a single cluster. Then we find the two clusters that are most similar in behavior and combine them into one cluster. We repeat the above steps until the number of preset clusters is reached. The following are the methods to calculate the similarity between two clusters. Firstly we choose one cluster as an initial cluster, compared to other clusters, we give the same apps transfer behavior negative weight. On the contrary, we give different apps transfer behavior positive weight. And finally the two clusters with the smallest sum of weights is combined to form a new cluster. For clusters with more than one data point, we choose the average-linkage method in the experiment, which means calculating the average of the distance between every two data points in the two clusters as the distance of the two clusters. After clustering, the overall characteristics will be represented by the characteristics of the cluster. Through the analysis of the feature clusters, the behavioral characteristics of the Internet in the two cities can be obtained.

Silhouette Coefficient is an evaluation method of clustering results. We can determine the number of preset clusters by solving the silhouette coefficients to achieve the best discrimination of the sample. The following is the calculation formula:

$$S_i=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \qquad (2)$$

In the formula, $a(i)$ represents the average value of the dissimilarity between the online behavior pattern of individual i and the online behavior pattern of other individuals in the same cluster, and the value is obtained according to the similarity calculation method in the cluster. The smaller the value is, the more the individual i should belong to the class. $b(i)$ represents the maximum value of the dissimilarity between the online behavior pattern of individual i and the online behavior pattern of other individuals in other clusters. The larger the value is, the less the individual i should belongs to the cluster. S_i represents the Silhouette Coefficient of the vector, and the mean of S_i for all samples is called the Silhouette Coefficient of the clustering result. The Silhouette Coefficient is used to determine the number of preset clusters, which makes the Silhouette Coefficient larger.

4) Tucker Decomposition verifies the result

Usually the result of clustering often has a certain degree of deviation, which make the method a little uncertain. So we adopt Tucker decomposition to verify it. Tucker Decomposition is a high-order principal component analysis representing a tensor as a core tensor multiplied by a matrix along each mode. Each user's behavioral characteristics can be expressed as a fixed weight of the nuclear tensor, so the user's behavior and weight are one-to-one correspondence. Almost all the existing algorithms for Tucker decompositions require certain processing based on the full tensor during the estimation [26]. The tensor is a multidimensional form of data storage, and the dimension of the data is called the tensor order. The following figure can represent the process of Turkey Decomposition.

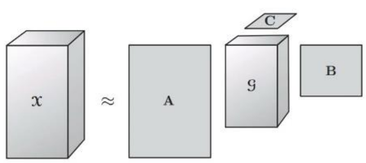


Fig.1 Tucker decomposition schematic

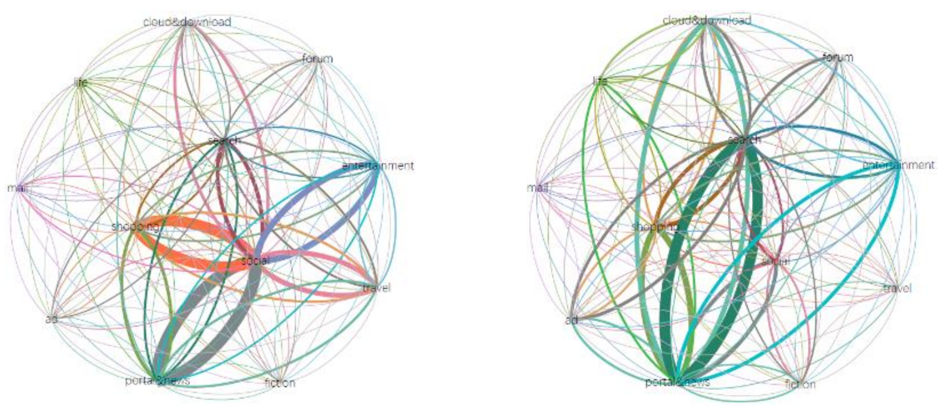
Its expression is:

$$
\begin{aligned}
X_{(1)} &\approx AG_{(1)}(C \otimes B)^T \\
X_{(2)} &\approx BG_{(2)}(C \otimes A)^T \\
X_{(3)} &\approx CG_{(3)}(B \otimes A)^T
\end{aligned}
\tag{3}
$$

In the formula, G represents the core in the decomposition process. In the experiment, we call it the behavioral core, which usually reflects a certain way of online behavior. A, B, and C are characterized by three dimensional coefficients of tensor, and the values of different dimensional coefficients will determine the weight of the core value. In the calculation process, the original online behavior mode is approximated by the weighting of the coefficient and the core as much as possible. Then we perform Tuck Decomposition on the sample. And the decomposed results are compared with the clustering results to verify the accuracy of the clustering results. We re-cluster if the deviation between the results is too large.

IV. Analysis of Experimental Results

According to the above methods, firstly, we integrate the states transition data of the individual by the MSM model, and basing on this, we draw the transfer frequency diagrams of various Apps as follows. Figure 2 shows the transfer information between apps used by netizens in Beijing and Jinhua City. In this figure, 12 nodes are represented as 12 different types of Apps, and the two-way connection line between different apps indicate that there is a transfer relationship between the two types of apps. Lines in different colors and of different thicknesses are respectively represented as different transfer relationships and different transfer intensities. The thicker the line is, the larger the number of transfers is. By observing the transfer frequency figure, we can intuitively obtain the data usage of various apps and the number of transfers between different types of Apps. So we can draw the following conclusions which are helpful for Internet advertising investment.



(a) (b)

Fig.2 Apps transfer frequency diagrams. (a): Apps transfer frequency diagram in Beijing (b): Apps transfer frequency diagram in Beijing

*1)*Through Figure 2, we can conclude that the social class apps have a very high frequency transfer. Therefore, such apps play a bridge role in the use of the mobile Internet, and they have obvious two-way connections with other types of apps. Thus, in order to gain a higher level of attention, we recommend that Internet advertising companies increase their budgets for social class apps when allocating investments.

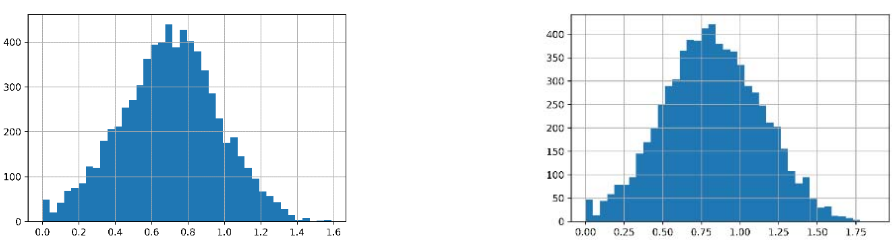
*2)*Generally speaking, entertainment type of apps is always time consuming. The longer time they consume, the greater business value they contain. However, by observing Figure 2, we can find that the transfer frequency of entertainment apps is not as good as that of social apps, which means the range of entertainment apps transfer is narrower. Therefore, when investing in the Entertainment class apps, advertisers need more targeted and reasonable advertising investment. Selection: Highlight all author and affiliation lines.

*3)*Both figures above show that reading, life, and mail apps have relatively low traffic. So in order to get better returns, investors should analyze costumers' needs accurately when advertising.

*4)*We can see that information class apps have a strong connection with entertainment and shopping class apps in subgraph (b). Therefore, the information apps can cause a kind of leisure and entertainment desire of netizens. Meantime information apps have a very large traffic base, which means it is an important point for advertisers.

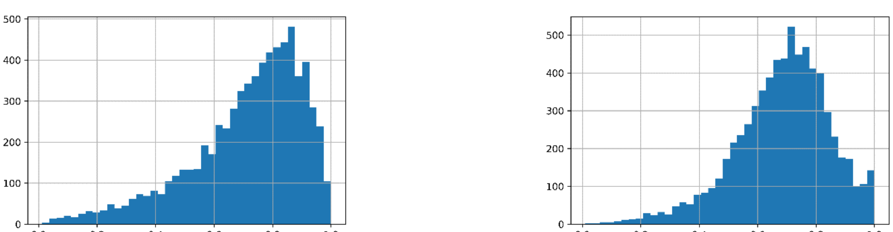
After we find the transfer relationship between apps, we try to link this relationship to the characteristics of the netizen. So we have obtained the conditional entropy and normalized conditional entropy plots of the two cities: figure 4 and figure 5. In those two figures, (a) (b) two subgraphs represent the entropy of Beijing and Jinhua, respectively. The larger the horizontal axis of their peaks is, the more dispersed the Internet users in the area are.

So from those two figures we can see that compared with Jinhua, Beijing has a small transfer entropy and normalized entropy. It can be seen that the behavior of the people in Jinhua with lower consumption levels is more diverse. And people in Beijing with higher consumption levels have more fixed online behavior patterns, which means their behavioral preferences are not easily influenced by advertising. Therefore, advertising may have greater benefits for people with lower levels of consumption



(a) (b)

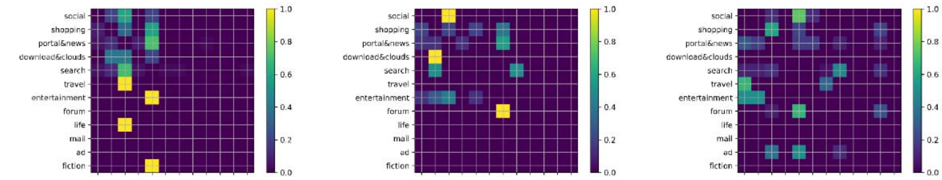
Fig.3 Conditional entropy of the state transfer model.(a): Conditional entropy of the state transfer model in Jinhua (b): Conditional entropy of the state transfer model in Beijing



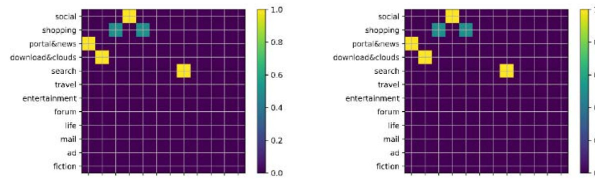
(a) (b)

Fig.4 Normalized conditional entropy of the state transfer model. (a): Represent Jinhua (b): Represent Beijing

In order to make the above results more intuitive and specific, we use hierarchical clustering and tucker decomposition. Hierarchical clustering divides the people of the two places into ten categories by combining people with similar behaviors. Figures 5 and 6 respectively show these ten classes behavioral characteristics. In those figure, the horizontal and vertical coordinates are expressed as the type of apps and the intersection of horizontal and vertical coordinates represents the transfer between two types of apps. The color depth of the junction points indicates their transition probability. It corresponds to the color comparison table on the right side of the figure. Figure 7 and 8 show the results of tensor decomposition of Jinhua and Beijing, respectively. The four subgraphs are represented as four cores, and all the behavior patterns in the sample are represented by the weight of the cores. We use it mainly to eliminate uncertainty of clustering results.

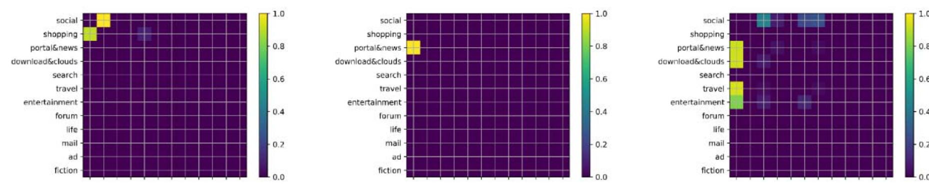


(a) (b) (c)

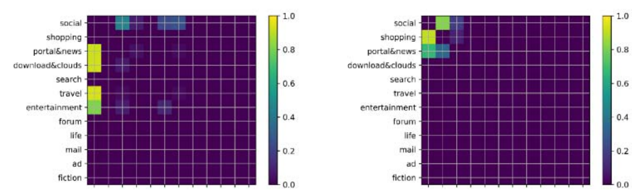


(d) (e)

Fig.5 Clustering results of Internet behavior of netizens in Jinhua City. (a)(b)(c)(d)(e): represent five different kinds of behavioral patterns of netizens in Jinhua



(a) (b) (c)



(d) (e)

Fig.6 Clustering results of Internet behavior of netizens in Beijing City.
(a)(b)(c)(d)(e): represent five different kinds of behavioral patterns of netizens in Jinhua

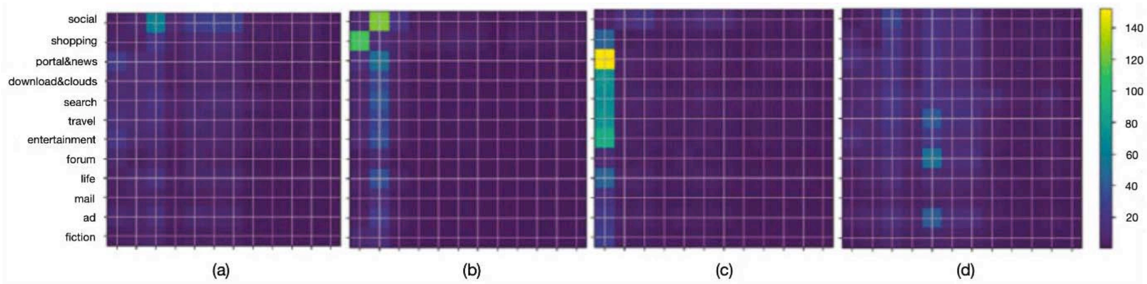


Fig.7 Beijing Tucker decomposition results

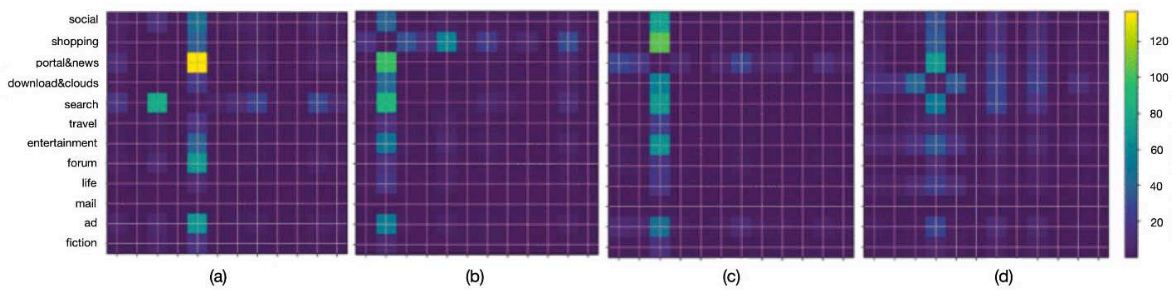


Fig.8 Jinhua City Tuck decomposition results

Observing the above figures, we get the following conclusion:

1) In figure 6, people use the social apps most widely, but differently, in figure 5, people regard the information as the medium. So it is more effective to put those ads targeted high-spending people on social apps, and information apps are more suitable for mass-market advertisers.
2) All of the figures show that after using communication apps, people will be interested in shopping, travel and entertainment. So these types of advertising investors should focus on social apps.
3) Only the two types of people of (a) (d) in Figure 6 are interested in other products when using life and fiction Apps. Therefore, the advertisers who invest in these two types of apps will get relatively less revenue.

V. Conclusions and Prospects

this paper analyzes the behavior orientation of people in different apps usage situations, and then solves the problem of untargeted delivery in the process of Internet advertising. We use the MSM model to build the transfer relationship of Internet users when they are online. Based on the transfer relationship of individual, we use Hierarchical Clustering to mine potential links between characteristics of the netizen and apps transfer, and then obtain more intuitive analysis results. Finally we use Tucker Decomposition to verify clustering results. After comprehensively analyzing the above results, we derive some methods to achieve accurate delivery of Internet advertising.

A. conclusion

The results show that in the competition of Internet advertising, advertisers targeting low-to-medium consumers will get more benefits. Social apps are suitable for advertisers in entertainment, travel and shopping. Information Apps are more suitable for mass-market advertisers. The Reading and Life apps have a small audience, and the user habits are relatively independent. It is not easy to generate ideas for deflection in other fields, and is not suitable for large-scale advertising.

B. Prospects

However, the research in this paper has the following problems in explaining the details of specific behavioral preferences. We doesn't take those similar apps into account which has different functions. Today apps have a wide variety of features, and an app may both have the function of social communication and reading. Therefore, the method of dividing the apps type by the main function has a certain deviation. At the same time, we do not consider the factor of the length of using time in the process of considering the transfer situation, which also has certain discrepancies with the real situation. These problems need to be resolved in next research, in order to more fully and accurately reveal the behavioral tendency of netizens in different network environment, which can provide a more reasonable and accurate reference for Internet advertising.

References

[1] Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. Proceedings of the VLDB Endowment, 5(12), 2032-2033.

[2] Lynch, C. (2008). Big data: How do your data grow?. Nature, 455(7209), 28..

[3] Data, data everywhere. The Economist (http://www.economist.com/node/15557443), Feb 2010.

[4] Chen,Y.,&Xie,J.(2005).Third-party product review and firm marketing strategy. Marketing Science, 24(2), 218-240.

[5] Lu, J. (2017). Study of informative advertising competition model in duopolistic market with relative profit object. Journal of Service Science and Management, 10(02), 105. 6.

[6] Edwards,J.(2005).Pharm formulates plans to move marketing from TV. Brandweek, 46.

[7] Cho,C.H.,&as-,U.O.T.A.A.I.A.(2004).Why do people avoid advertising on the internet?. Journal of advertising, 33(4), 89-97..

[8] Levinson, P. (2004). Cellphone: The story of the world's most mobile medium and how it has transformed everything!. Macmillan.

[9] Qiu, M., Jia, Z., Xue, C., Shao, Z., & Sha, E. H. M. (2007). Voltage assignment with guaranteed probability satisfying timing constraint for real-time multiproceesor DSP. The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology, 46(1), 55-73.

[10] Qiu, M., Sha, E. H. M., Liu, M., Lin, M., Hua, S., & Yang, L. T. (2008). Energy minimization with loop fusion and multi-functional-unit scheduling for multidimensional DSP. Journal of Parallel and Distributed Computing, 68(4), 443-455.

[11] Krum, C. (2010). Mobile marketing: Finding your customers no matter where they are. Pearson Education.

[12] Martin, C. (2011). The third screen: marketing to your customers in a world gone mobile.

[13] Yu Kun. (2012). Research on user acceptance model of mobile internet advertising based on UTAUT model (Doctoral dissertation, Beijing: Beijing University of Posts and Telecommunications).

[14] Yu Ying, & Liu Wenjun. (2010). Media, Scene, Behavior - Talking about Merowitz's Mediaal Dissertation from The Lost Region.

[15] Johnson, E. J. (2001). Digitizing consumer research. Journal of consumer Research, 28(2), 331-336.

[16] Sun Haiyang (2017). Design and application of scene interaction in mobile consumption [D]. Jiangnan University.

[17] Wang, J., Qiu, M., & Guo, B. (2017). Enabling real-time information service on telehealth system over cloud-based big data platform. Journal of Systems Architecture, 72, 69-79.

[18] Chen, M., Zhang, Y., Qiu, M., Guizani, N., & Hao, Y. (2018). SPHA: Smart personal health advisor based on deep analytics. IEEE Communications Magazine, 56(3), 164-169.

[19] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.

[20] Tian, T., Zhu, J., Xia, F., Zhuang, X., & Zhang, T. (2015, May). Crowd fraud detection in internet advertising. In Proceedings of the 24th International Conference on World Wide Web (pp. 1100-1110). International World Wide Web Conferences Steering Committee.

[21] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.

[22] Schein, A., Zhou, M., Blei, D. M., & Wallach, H. (2016). Bayesian poisson tucker decomposition for learning the structure of international relations. arXiv preprint arXiv:1606.01855.

[23] Jackson, C., & Jackson, M. C. (2018). Package 'msm'.

[24] Collins, D. M. (1982). Nature (London), 298, 49–51

[25] Johnson, S. C. (1967). Hierarchical clustering schemes. Psychometrika, 32(3), 241-254.

[26] Renard, N., & Bourennane, S. (2009). Dimensionality reduction based on tensor modeling for classification methods. IEEE Transactions on Geoscience and Remote Sensing, 47(4), 1123-1131.