

## Recognition of the Purchasing Intentions of WeChat Users Based on Forgetting Curve

Feng Yang<sup>1</sup>, Baoxin Liu<sup>1\*</sup>, Laiqin Zhao<sup>1</sup>, Xiangfang Peng<sup>2</sup>

<sup>1</sup>School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China

<sup>2</sup>Linyi Branch of China Mobile Communications Group Shandong Co. LTD, Linyi 276000, China

Corresponding Author Email: [inspurliu@163.com](mailto:inspurliu@163.com)

<https://doi.org/10.18280/ria.330111>

**Received:** 23 December 2018

**Accepted:** 5 February 2019

### **Keywords:**

*intention recognition, forgetting curve, WeChat, data mining, big data, prediction, purchasing intention*

### **ABSTRACT**

The chat data of WeChat are particularly important in the era of big data. The recognition of the purchasing intentions of WeChat users can enable sellers to make accurate recommendations. However, the users' interests in items are every-changing, making it hard to accurately identify their purchasing intentions. To solve the problem, this paper puts forward a purchasing intention recognition model based on the forgetting curve, a key tool in the experimental psychology of memory. The model predicts the current purchasing intentions of a user based on his/her historical information, and judges the strength of the intentions according to the time of the historical information. If a user is interested in an item, he/she tends to mention the item repeatedly and has a deep impression of the item. This tendency is similar to the laws of forgetting and repeated learning. The similarity ensures the accuracy of the proposed model in recognition of users' purchasing intentions. The model was proved experimentally to have high recognition accuracy. The research findings deepen the understanding of users' historical information.

## 1. INTRODUCTION

In the era of big data, countless amounts of data are generated on a daily basis. This calls for efficient techniques of data mining. As the leading social networking app in China, WeChat boasts a staggering number of over 300 million users, creating a gigantic amount of data in their chats. These chat data are a wealth of personal and interactive information. Through rational and effective mining and analysis, it is of great commercial value to derive from these data the purchasing habits and intentions of users.

The recognition of purchasing intentions faces multiple challenges, such varied expression of texts and the difficulty in accurate understanding of user intentions [1]. The traditional approaches of intention recognition are grounded on template matching methods or feature-based learning methods. However, these methods are too limited to pinpoint the user intentions, failing to consider the impacts of varied intentions on the work. In modern society, the user interests often change with the elapse of time. For instance, the user interests in an item may shift to another item in a period of time. This phenomenon is known as the interest drift [2].

Interest drift is commonplace in the recognition of purchasing intentions. The impacts of interest drift must be taken into account before recognizing the purchasing intentions of a user. In light of the above, this paper puts forward a recognition method for purchasing intentions based on the forgetting curve, a key tool in the experimental psychology of memory, aiming to disclose the impacts of interest drift on recognition.

## 2. LITERATURE REVIEW

To recognize the purchasing intentions of WeChat users, the key is to determine users' intentions through analysis on the chat data. There is little report on the recognition of the purchasing intentions among WeChat users, because the research on WeChat data mining is still in its infancy. Nevertheless, the economic benefits of WeChat-based intention recognition have attracted wide attention from the industry and academia [3]. For instance, the intention recognition helps the sellers understand the users' demand and make accurate recommendations to the users.

Through its research on major industries in the US, Double Click Inc. Double Click Search found that most consumers tend to search for and discuss about the items they want to buy on the Internet before purchasing these items (Double Click Search before the Purchase-Understanding Buyer Search Activity as it Builds to Online Purchase: [http://www.doubleclick.com/insight/pdfs/searchpurchase\\_0502.pdf](http://www.doubleclick.com/insight/pdfs/searchpurchase_0502.pdf)). Goldberg *et al.* [4] proposed the concept of "buy wish", a.k.a. the purchase intention, and acquired the purchase intentions of users from their shopping lists and comments on the items. Zhang and Nasraoui [5] combined the query flow graph and the text into a new query recommendation algorithm. Kooti *et al.* [6] suggested that people's consumption behaviour is affected by age, gender and other factors.

With the boom of social networking and chat apps in recent years, WeChat, Weibo and Twitter have become the main tools for young people to release and exchange information. Yang and Li [7] collected the product information of people's daily necessities and their hashtags in Twitter, and recognized users' purchasing intentions through template matching. Considering the huge amount of data and the lack of tags, Liu

et al. [8] designed a sorting algorithm based on weakly supervised graphs for intention recognition.

Unlike the traditional ways to recognize purchasing intentions, this paper conducted recognition of the chat data of WeChat in view of the popularity of new media among the masses. Meanwhile, the forgetting curve was adopted to recognize the purchasing intentions revealed in the chats between WeChat users, considering the impacts of interest drift [9].

### 3. PURCHASING INTENTION ANALYSIS OF WECHAT USERS

#### 3.1 Correlation between WeChat chat data and users' purchasing intentions

The chat data of WeChat are highly fragmented and scattered among users and terminals. As a result, it is impossible to analyse the unstructured data, not to mention remembering the communication scene and identity of the users. Despite the difficulties, the potential value of WeChat chat data is worth exploring, because data are the core of setting up competitive barriers. Hence, this paper attempts to analyse the chat data of WeChat and disclose the purchase intentions revealed in the chats between WeChat users.

In WeChat, the most common and basic behaviour is the spontaneous chats between users. These user-initiated chats are proactive in nature. On the macroscale, the chat contents can be regarded, to a certain extent, as the subjective expression of users and the mirror of users' interests. In a narrow sense, however, the sentences containing users' purchasing intentions are hard to decipher, due to the diversified and individualized expressions of the chat contents [10].

#### 3.2 Purchase intentions and forgetting curve

In this paper, the author attempted to determine the products currently favoured by users in light of their chat history. The users pay attention to different topics over time. The originally favourite topics gradually fade away and give their prominence to new topics. In other words, the value of historical information is negatively correlated with its distance to the current time. This trend bears high resemblance with the law of forgetting things over time, and can thus be depicted by the forgetting curve. Meanwhile, an attractive topic tends to be mentioned repeatedly, similar to how important knowledge points are repeatedly reviewed in the learning process. The forgetting curve is constantly adjusted in the repeated learning process.

Considering the above, the forgetting curve was added to the model of intention recognition, creating a new way to recognize the purchasing intentions based on the forgetting curve.

### 4. RECOGNITION OF PURCHASE INTENTIONS OF WECHAT USERS BASED ON FORGETTING CURVE

#### 4.1 Forgetting curve model

The German psychologist Ebbinghaus was the first to study human forgetting law through scientific experiments. The

natural forgetting curve of human beings is named as the Ebbinghaus forgetting curve after him. The forgetting law is shown in Table 1 below.

**Table 1.** The forgetting law

Elapsed time since learning	Retention (%)
Immediately	100
20 minutes	58
1 hour	44
9 hours	36
1 day	33
2 day	28
6 day	25
31 day	21

Based on the Ebbinghaus forgetting curve, quantitative analysis shows that the amount of memory changes with time: the attenuation is fast at the beginning and gradually tends towards a stable rate; after a period of time, a part of the memory is forgotten, and the rest is saved in mind as the memory value.

Zeng et al. [11] mathematically analysed the forgetting law, and fitted the law of human forgetting curve with a negative exponential curve. The resulting function expression is:

$$p(t, k) = p_0 e^{-kt} \quad (t > 0) \quad (1)$$

where,  $k$  is the attenuation rate;  $p_0$  is the initial memory value.

When users chat frequently in a short time, their purchasing intention of an item is considered as attenuating at a constant rate, i.e. the attenuation of the intention conforms to the attenuation law of a single memory curve. Hence, the strength of user  $a$ 's purchasing intention of item  $p_i$  at time  $t_n$  can be expressed as:

$$S_a(t_n, p_i) = S_a(t_0, p_i) * e^{-k_a t_n} + \vec{K}_{t_n} \cdot \vec{P}_{t_n} \cdot e^{-n} \quad (t_n > t_0) \quad (2)$$

$$S_a(t_0, p_i) = \vec{K}_{t_0} \cdot \vec{P}_{t_0} \quad (3)$$

Equation (3) shows the user  $a$ 's purchasing intention of item  $p_i$  at time  $t_0$ , i.e. in the initial phase. Let  $K$  be the set of the user's keywords extracted at time  $t_0$  and  $P$  be the set of items the user is interested in. Then, the user's purchasing strength at time  $t_0$  can be obtained based on  $K$  and  $P$ . The user's purchasing strength at time  $t_n$  can be determined through the same steps. In addition, let  $k_a$  be the forgetting rate of user  $a$ . The forgetting rate differs from user to user, because of the user difference in repeated learning. A user may be interested in an item at some point due to a stimulant, but the interest level will decline over time. If the same stimulant occurs, user  $a$ 's interest in the item will gradually decrease and approach a constant value. Here, the incremental stimulation of a keyword to the purchasing intention of an item is adjusted via  $e^{-n}$  under the stimulant of  $\vec{K} \cdot \vec{P}$ . Note that  $n \in (0, 1, 2, 3, \dots) = N-1$ , with  $N$  being the number of cumulative time points.

#### 4.2 Multi-stage quantification of intention strength

The forgetting curve can reflect the change of human memory value, but only that in a single time period. It cannot

be directly applied to describe users' purchasing intentions revealed in chats. If a user is desirous of purchasing an item, he/she will talk about many topics on the items continuously. In this case, the chat is frequent in the short term and continuous in the long term. In another case, an item is not mentioned throughout the chat because it does not attract the user's interest. However, the item might be mentioned suddenly on another day. Thus, a multi-stage quantification method was proposed for intention strength by adding these situations into the forgetting curve [12].

The user's chat on the same topic at different times is equivalent to the repeated learning of an interest. Thus, the moment of repeated learning is the time when the keyword is mentioned again. Let  $t_1$ ,  $t_2$  and  $t_3$  be the moments of repeated learning. Then, the change of user intention can be illustrated as Figure 1 below.

As shown in Figure 1, the repeated learning can be divided into multiple stages. Each stage represents a new process of natural forgetting, and has a unique forgetting rate and initial memory value. To measure the real-time purchasing intention, it is necessary to determine the initial memory value  $S(a,n)$  and forgetting rate  $k(a,n)$  of each stage.

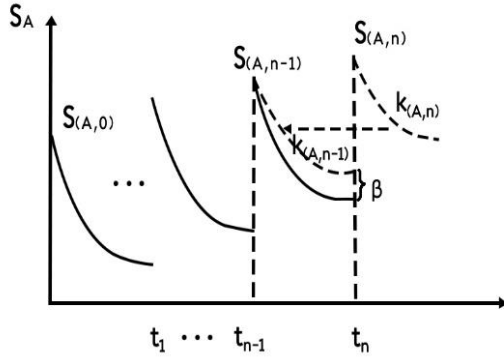


Figure 1. The change of user intention

It can be seen that the initial memory value  $S(a,n)$  of stage  $n+1$  is higher than the final memory value of stage  $n-1$ . The difference between the two values is referred to the repeated learning increment. The initial memory value of stage  $n+1$  can be expressed as:

$$S(a,n) = S(a,n-1) * e^{-k_{(a,n-1)}(t_n - t_{n-1})} + \vec{K}_{t_n} \cdot \vec{P}_{t_n} \cdot e^{-n} \quad (4)$$

It can be seen from Figure 1 that the curve of forgetting rate  $k(a,n)$  can be overlapped with the curve of forgetting rate  $k(a,n-1)$  after moving to the lower left. Therefore, the relationship between the two curves can be obtained from that between  $k(a,n)$  and  $k(a,n-1)$ . The height difference  $\beta$  between the two curves can be determined by moving the initial point of the curve for stage  $n+1$  to that of the curve for stage  $n$  in Figure 1. The value of  $\beta$  can be calculated as:

$$\beta = S(a,n-1) \left[ e^{-k_{(a,n)}(t_n - t_{n-1})} - e^{-k_{(a,n-1)}(t_n - t_{n-1})} \right] \quad (5)$$

From the nature of the exponential function, the maximum value of  $\beta$  can be determined as  $\beta = S(a,n-1) [1 - e^{-k_{(a,n-1)}(t_n - t_{n-1})}]$ . Assuming that  $\eta = t_n - t_{n-1}$ , we have:

$$k(a,n) = \frac{\ln \left( \frac{\beta}{S_{(a,n-1)}} + e^{-k_{(a,n-1)}\eta} \right)}{-\eta} \quad (6)$$

The above equation shows that the value of  $k(a,n)$  can be determined based on the value of  $\beta$ . In addition, the adjustment degree of the forgetting rate in each repeated learning, denoted as  $\theta$ , should also be known. The value of  $\theta$  is positively correlated with the time gap between two adjacent repeated learning processes, i.e. the interval between a user's mentioning of the same item. Here,  $\theta$  is called the inertia factor, because its value reflects how lazy a user is to mention an item. Then, the maximum value of  $\beta$  can be divided into  $\theta$  segments:

$$\beta = \frac{S(a,n-1) [1 - e^{-k_{(a,n-1)}(t_n - t_{n-1})}]}{\theta} \quad (7)$$

Substituting equation (7) into equation (6), we have the expression for  $k(a,n)$ :

$$k(a,n) = \frac{\ln [1 + (\theta - 1)e^{-k_{(a,n-1)}\eta}] - \ln \theta}{-\eta} \quad (8)$$

The value of  $k(a,n)$  can be determined when the following parameters are known, including the forgetting rate  $k(a,n-1)$ , the time segment  $\eta$  and the inertia factor  $\theta$ . Then, the strength of the purchasing intention of each stage can be obtained.

## 5. EXPERIMENT AND RESULTS

### 5.1 Preparations

(1) Experimental data. The experimental data were extracted from the WeChat Open Platform (<https://open.weixin.qq.com/>). The original data contain the chat and personal information of 200 plus users from June to November, 2016. After data processing, a total of 1,353,564 data entries were obtained for our experiment.

(2) Parameter settings. The experimental parameters include the initial forgetting rate  $k_0$ , the initial memory value of the first stage  $S_0$  and the base of the inertia factor  $\theta_0$ . The parameter settings are listed in Table 2 below.

Table 2. Parameter setting

Name	Meaning	Value
$K_0$	Initial forgetting rate	1.0
$S_0$	Initial memory value of the first stage	1.0
$\theta_0$	Base of the inertia factor	8

### 5.2 Evaluation indices

The correctness of keyword extraction was evaluated against the recall rate and accuracy:

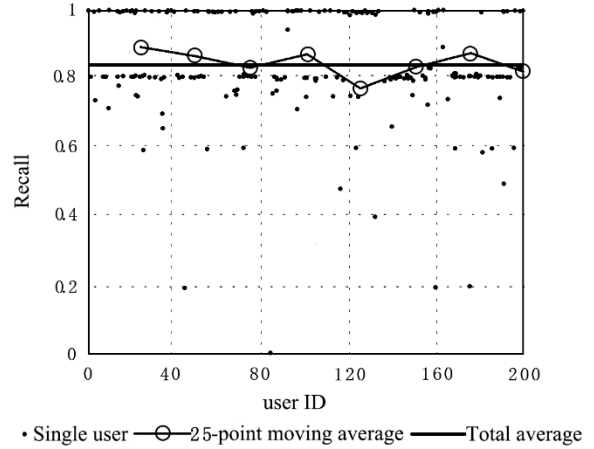
$$R = \frac{TP}{TP + TN} \quad (9)$$

$$P = \frac{TP}{TP + FN} \quad (10)$$

where,  $TP$  is the number of extracted keywords that are consistent with the correct keywords;  $TN$  is the number of extracted keywords that are inconsistent with the correct keywords;  $FN$  is the number of extracted keywords that are not the correct keywords.

### 5.3 Experimental results and analysis

(1) Model accuracy. The recall rates of our model is shown in Figure 2, where each solid point is the recall rate of each user, each hollow point is the mean recall rate of every 25 users, and the horizontal line is the mean recall rate of all 200 users. It can be seen from the figure that the recall rate of over 42% of the users was one, that of over 85% of the users was greater than 0.70, and the mean recall rate of all users was about 0.847. The results indicate the our model can basically rank the items favoured by the users in a correct manner [13].



**Figure 2.** The recall rates of our model

(2) Contrastive experiment. The proposed intention recognition strategy was further contrasted with the sentiment analysis of Baidu AI Open Platform (<http://ai.baidu.com/>). The results of two random users are displayed in Table 3 below.

**Table 3.** The results of two random users

	Method	User A	User B
Sentiment analysis		(Hydrating mask, positive)	(Acne mask, negative)
		(Whitening mask, negative)	(Hydrating mask, positive)
		(Moisture mask, positive)	(Whitening mask, positive)
	...	...	...
Purchasing intention recognition based on forgetting curve		(Hydrating mask, 74%)	(Acne mask, 7%)
		(Whitening mask, 13%)	(Hydrating mask 76%)
		(Moisture mask, 69%)	(Whitening mask, 83%)
	...	...	...

Table 3 gives the prediction results on user a and user b. According to the sentiment analysis, user a is positive towards hydrating mask and moisture mask, and negative towards whitening mask; user b is positive towards hydrating mask and whitening mask, and negative towards acne mask. The proposed model predicts that the strength of user a’s purchasing intention is 74% for hydrating mask and only 13% for whitening mask, while the strength of user b’s purchasing intention is 76% for hydrating mask and only 7% for acne mask. Through comparison, it is learned that the sentiment analysis can only judge if a user is positive or negative towards an item, while the proposed model can predict the strength of a user’s purchasing intention for an item. With our model, the seller can identify the most favoured items among users and make accurate recommendations [14].

(3) Results analysis. Among the traditional recognition methods for users’ purchasing intentions, the template matching method only considers the number of occurrences of words and features in the text. This method cannot adapt to the diverse chat texts or the complex situations of WeChat sessions, failing to achieve a desirable recall rate.

Another popular recognition method is the local text perception based on convolutional neural network (CNN), which obtains eigenvectors through convolution with multiple kernels and retains the important textual information by pooling these eigenvectors. Considering the deep information of the texture, this recognition method can excavate the deep meaning from the text. However, the recall rate of this method is not satisfactory, because the context is lost in the pooling

operation, making it impossible to fully understand the semantics based on the context.

The above two methods have different degrees of limitations. The most severe defect of them lies in the overlook or underestimation of users’ historical information. In this paper, the chat texts are assigned different values according to the law of the forgetting curve. The value of a chat text is negatively correlated with its distance to the current information. In addition, the value of historical information is re-adjusted in repeated learning through the multi-stage quantification of intension strength. In this way, our model fully considers the changes in purchasing intentions resulted from the interest drift, and works excellently in the experimental recognition of purchasing intentions.

## 6. CONCLUSIONS

With the rapid growth in the number of WeChat users, it is very meaningful to recommend items based on the users’ purchasing intentions. The recognition of purchasing intentions is a research hotspot. However, the common methods like feature-based machine learning and template matching fail to consider the impacts of interest drift, or achieve a high accuracy. To solve the problem, this paper puts forward a new way to recognize purchasing intentions based on the forgetting curve, in light of the users’ historical information and the impacts of interest drift. The model was proved experimentally to have high recognition accuracy. The

research findings deepen the understanding of users' historical information.

## REFERENCES

- [1] Fu, B., Liu, T. (2015). Consumption intent recognition for social media: Task, challenge and opportunity. *Intelligent Computer and Applications*, (4): 1-4. [Http://dx.doi.org/10.3204/DESY-PROC-2008-02/laarmann-tim](http://dx.doi.org/10.3204/DESY-PROC-2008-02/laarmann-tim)
- [2] Dai, H.K., Zhao, L., Nie, Z., Wen, J.R., Wang, L., Li, Y. (2006). Detecting online commercial intention (OCI). In: *Proceedings of the 15th International Conference on World Wide Web*, Edinburgh, pp. 829-837. <https://doi.org/10.1145/1135777.1135902>
- [3] Zhang, F., Yuan, N.J., Lian, D., Xie, X. (2014). Mining novelty-seeking trait across heterogeneous domains. In: *Proceedings of the 23rd International Conference on World Wide Web*, Seoul, pp. 373-384. <https://doi.org/10.1145/2566486.2567976>
- [4] Goldberg, A.B., Fillmore, N., Andrzejewski, D., Xu, Z., Gibson, B., Zhu, X. (2009). May all your wishes come true: a study of wishes and how to recognize them. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Boulder, pp. 263-271.
- [5] Zhang, Z., Nasraoui, O. (2008). Mining search engine query logs for social filtering-based query recommendation. *Applied Soft Computing*, 8(4): 1326-1334. <https://doi.org/10.1016/j.asoc.2007.11.004>
- [6] Kooti, F., Lerman, K., Aiello, L.M., Grbovic, M., Djuric, N., Radosavljevic, V. (2015). Portrait of an online shopper: Understanding and predicting consumer behavior. *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*, Insead, pp. 205-214. <https://doi.org/10.1145/2835776.2835831>
- [7] Yang, H., Li, Y. (2013). Identifying user needs from social media. *IBM Res Div*, 1-10.
- [8] Liu, T., Fu, B., Chen, Y.H. (2015). Detecting consumption intention based on graph ranking in social media. *Scientia Sinica (Informationis)*, 45: 1523-1535.
- [9] Cui, L., Shi, Y. (2014). A method based on one-class SVM for news recommendation. *Procedia Computer Science*, 31: 281-290. <https://doi.org/10.1016/j.procs.2014.05.270>
- [10] He, J.C. (2012). *Research on frequency pre-equalization for MIMO broadcast wireless channel*. Xi'an: Xidian University.
- [11] Zeng, D., Wang, T., Yan, S., Lai, H. (2013). A collaborative filtering algorithm based on exponential forgetting function. *Science Mosaic*, (7): 10-15.
- [12] Donahue, J., Hendricks, L.A., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko, K., Darrell, T. (2015). Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, London, pp. 2625-2634.
- [13] Hao, H.R. (2016). New mixed kernel functions of SVM used in pattern recognition. *Cybernetics and Information Technologies*, 16(5): 5-14. <http://dx.doi.org/10.1515/cait-2016-0047>
- [14] Sainath, T.N., Vinyals, O., Senior, A., Sak, H. (2015). Convolutional, long short-term memory, fully connected deep neural networks. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, London, pp. 4580-4584. <http://dx.doi.org/10.1109/ICASSP.2015.7178838>