A RNN-Based Multi-factors Model for Repeat Consumption Prediction

Zengwei Zheng[1], Yanzhen Zhou[1,2], Lin Sun[1(✉)], and Jianping Cai[1]

[1] Hangzhou Key Laboratory for IoT Technology and Application, Zhejiang University City College, Hangzhou, China
{zhengzw,sunl,jpcai}@zucc.edu.com

[2] College of Computer Science and Technology, Zhejiang University, Hangzhou, China
zhouyanzhen@zju.edu.com

Abstract. Consumption is a common activity in people's daily life, and some reports show that repeat consumption even accounts for a greater portion of people's observed activities compared with novelty-seeking consumption. Therefore, modeling repeat consumption is a very important study to understand human behavior. In this paper, we proposed a multi-factors RNN (MF-RNN) model to predict the users' repeat consumption behavior. We analysed some factors which can influence customers' daily repeat consumption and introduced those factor in MF-RNN model to predict the users' repeat consumption behavior. An empirical study on real-world data sets shows encouraging results on our approach. In the real-world dataset, the MF-RNN gets good prediction performance, better than Most Frequent, HMM, Recency, DYRC and LSTM methods. We compared the effect of different factors on the customers' repeat consumption behavior, and found that the MF-RNN gets better performance than non-factor RNN. Besides, we analyzed the differences in consumption behaviors between different cities and different regions in China.

Keywords: Repeat consumption · Recurrent Neural Network (RNN) Multi-factors

1 Introduction

Nowadays, with the rapid development of mobile payment technology, people can make payment in an store by smartphones apps (such as Alipay, WeChat pay and Apple pay etc.) instead of by cash. Therefore, how to use previous consumption record and model user's repeat consumption behavior to predict which store the user likely to go in future time is very important. The study of consumption behavior is to know the way an individual spends his resources in the process of consuming items. This is an approach that comprises of studies of the items that they buy and the reason for buying and the timing. It is also about where

V. Kůrková et al. (Eds.): ICANN 2018, LNCS 11141, pp. 105–115, 2018.
https://doi.org/10.1007/978-3-030-01424-7_11

they make the purchase and how frequently. Due to the fact that people prefer the things they are familiar with, they may repeatedly interact with same items overtime. Therefore, users always like to visit the same stores that they have purchased previously, such as shopping at a same fruit shop and eating regularly at a same restaurant. For the reason of that, repeat consumption accounts for a major portion of people's daily consumption behavior, and we focus on the repeat consumption behavior study in this paper.

In real life, some factors can affect peoples' daily activities. For instance, we usually visit some places in the vicinity of our office during the workdays, but usually visit some places near our home in holidays; we probably go out for some outdoor activities when the weather is nice but stay indoor when the weather is terrible; we like to take cool drinks in summer day but choose some hot things in winter instead. Therefore, we believe that peoples' daily repeat consumption behavior can be affected by some factors too. In this paper, we proposed an MF-RNN model which is based on RNN and introduce some factors to predict peoples' repeat consumption behavior. Through analysis, we selected three factors as influential factors: holiday factor, weather factor and temperature factor. An empirical study on real-world dataset shows encouraging results on our approach. The MF-RNN gets encouraging performance for repeat consumption behavior prediction, better than MF, HMM, Recency, DYRC and LSTM model. And the MF-RNN with all three factors gets better performance than without any factors.

2 Related Works

Consumption behavior is an approach to know the way that an individual spends his resources in the process of consuming items. Consumption behavior analysis is critically extending the domain of behavior analysis and behavioral economics into marketing theory. In past, the ways of predicting the consumers' behavior involved Content-based recommendation, collaborative filtering-based recommendation, time series analysis and data mining. Content-based recommender systems are based on the idea that the features of items are useful in suggesting relevant and interesting items for users [1]. Collaborative filtering-based recommender systems identify users whose tastes are similar to that of a target user and then recommend items that the others have liked [2,3]. Time series analysis and data mining method used the historical data to extra some feature to model the user's consumption behavior [4].

But the study of repeat consumption behavior is a bit different of consumption behavior, it focuses on predicting whether or not the user will repeat purchase items which he has consumed in previous time. The problems of how and why users repeatedly consume certain items have been approached from several angles in various discipline [5]. Some of the earliest works focus on understanding repeat behavior on the web, like re-searching queries and website revisitation. Adar, Teevan and Fumais [6] carried out a large-scale analysis of revisitation, and classified websites into different groups based in how often they attract revisitors. Then, those researchers explored the relationship between the content change in web pages and people's revisition to these pages [7]. Teevan et al. [8] studied query logs to find repeat queries in web research, and that more than 40% of the queries are repeat queries. Then, many methods have been proposed to predict people's repeated consumption behavior. Anderson et al. [9] analyzed the dynamics of repeat consumption. They studied the pattern by which a user consumes the same item repeatedly over time, in some wide variety domains ranging from check-ins at the same business location to re-watches of the same video, and found that recency of consumption is the strong predictor consumption. Chen et al. [10] formulate the problem of recommendation for repeat consumption with user implicit feedback, then proposed a time-sensitive personalized pairwise ranking (TS-PPR) method based on user behavioral features. Rafailidis and Nanopoulos [11] present the CTF model and W-CTF model for recommend items with repeat consumption, by capturing the rate with which the current preferences of each user shift over time and by exploiting side information in a coupled tensor factorization technique. Zhang et al. [12] proposes a dining recommender system termed NDRS, which gives associated recommendation strategies according to different novelty-seeking statuses. They first designed a CRF (Conditional Random Field) with constrains to infer novelty-seeking status, then proposed a context-aware collaborative filtering method and a HMM (Hidden Markov Model) with temporal regularity method are proposed for novel and regular restaurant recommendation. Christina and Lars [13] developed the multinomial SVM (Support Vector Machine) item recommender system MN-SVM-IR to calculate personalized item recommendation for a repeat-buying scenario. Although there are many methods for predicting repeated consumption behavior, most methods focus on the features of the consumers or the items and rarely care about other relevant informations.

3 Methodolody

Recurrent Neural Network (RNN) is a type of feedforward neural network whose output is not only depend on the weight of the current input, but also depend on the present state of the network. Augmented by the inclusion of recurrent edges that span adjacent time steps, the RNN introducing a notion of time to the model [14]. In other words, the feedback from the hidden layer not only goes to the output, but also goes into the next time step hidden layer. Thus, the RNN has some memory. In the previous research, RNN proved to be very useful in sequence learning problem. RNN can be employed in text processing, image captioning, machine translation, video captioning and handwriting recognition. In this paper, we proposed a prediction model based on RNN and combine with several other influential factors to predict the users' repeat consumption behavior.

3.1 RNN-based Multi-factors Prediction Model

As mentioned above, we selected 3 different factors as the influential factors in the repeat consumption behavior prediction case. Then, we defined the MF-RNN model which is a three-layers network include input layer, hidden layer and output layer, shown in Fig. 1.

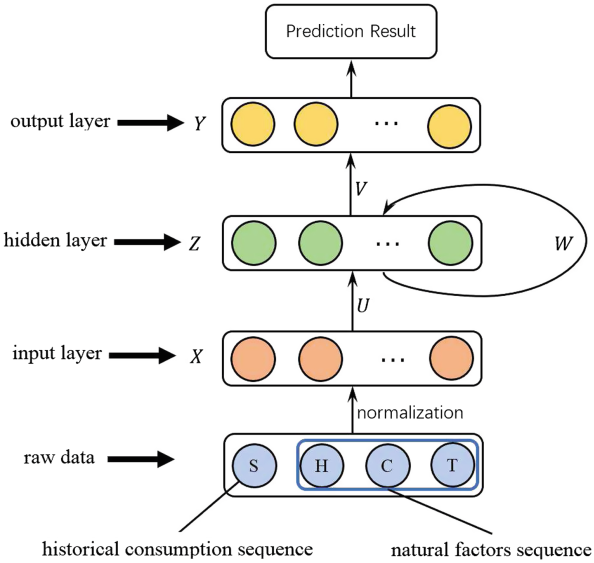

Fig. 1. The framework of MF-RNN.

Input Layer: The input layer X is a vector consists of four normalized input data as Eq. (1): S is visited offline store sequence, H is the holiday factor sequence, C is the weather factor sequence, T is the temperature factor sequence. The output data Y is the prediction result represent the offline store which this customer will visit in the next time.

$$X = [SHCT] \tag{1}$$

Hidden Layer: Z is hidden layer, its state in time t z_t is affected by the current input x_t and the state of the previous time step hidden layer z_{t-1}:

$$z_t = f(Ux_t + Wz_{t-1} + b_z) \tag{2}$$

where U is the weight between the input and hidden layers, W is the recurrent weight between the hidden layers at adjacent time steps, b_z is the bias in hidden layer.

Output Layer: The output layer Y is the prediction result represent offline stores the user will visit at next times. The output in time t calculate as Eq. (3), and V is the weight between the output and hidden layers, g is an activation function and b_y is the bias in output layer.

$$y_t = g(Vz_t + b_y) \tag{3}$$

Then, we used the historical data to training this network. A Back Propagation Through Time (BPTT) algorithm is employed in the training process to calculate the parameters U V W and b_z b_y. The loss function of the networks defined as Eq. (4), and e_t is the loss at each step.

$$E = \sum_t e_t \tag{4}$$

The gradient of V calculate as Eq. (5), y_t' is the supervision information at time step t.

$$\nabla V = \frac{\partial E}{\partial V} = \sum_t (y_t - y_t') \otimes z_t \tag{5}$$

Then we defined two operator δ_t^z as Eq. (6) and δ_t^y as Eq. (7), and calculate the gradient of U, W as Eqs. (8) and (9), finally calculate the gradient of two bias as Eqs. (10) and (11). After parameters training process, we got a trained network to calculate the output data Y, then to predict user's repeat consumption behavior in future time.

$$\delta_t^z = \frac{\partial E}{\partial (f(Ux_t + Wz_{t-1} + b_z))} \tag{6}$$

$$\delta_t^y = \frac{\partial E}{\partial (g\,(Vz_t + b_y))} \tag{7}$$

$$\nabla U = \frac{\partial E}{\partial U} = \sum_t \frac{\partial e_t}{\partial U} = \sum_t \delta_t^z \times x_t \tag{8}$$

$$\nabla W = \frac{\partial E}{\partial W} = \sum_t \frac{\partial e_t}{\partial W} = \sum_t \delta_t^z \times z_{t-1} \tag{9}$$

$$\Delta b_z = \frac{\partial E}{\partial b_z} = \sum_t \frac{\partial e_t}{\partial b_z} = \sum_t \delta_t^z \tag{10}$$

$$\Delta b_y = \frac{\partial E}{\partial b_y} = \sum_t \frac{\partial e_t}{\partial b_y} = \sum_t \delta_t^y \tag{11}$$

3.2 Influential Factors Selection

Holiday Factor: There is a big difference between peoples' daily activities on holidays and on workdays. For example, people prefer to choose the restaurant near their office room to have lunch on workdays but choose the restaurant near home to have lunch on holidays, and people can often visit supermarket during holidays but can't do it when they are at work. Therefore, the holiday factor will affect peoples' repeat consumption behavior. In this paper, according to Chinese statutory holiday arrangements, we generate a holiday sequence for each user, 1 represent holiday and 0 represent workday.

Weather Factor: Weather can also affect peoples' daily activities. For instance, people can do some outdoor activities like go to a playground and visit the park when the weather is good, but they always stay indoors when rainy and snowy. Then, we choose the weather as an influential factor in this repeat consumption study. The types of weather are diverse, including sunny, cloudy, rainy, snowy and etc. In this paper, we classify the weather into the following categories based on the types and the severity of the weather, and give them different labels, as shown in Table 1.

Temperature Factor: In addition to holiday factor and weather factor, temperature can also influence peoples' daily consumption behavior. When the temperature is very high, people may buy some cool drink or ice-cream. And when the temperature is low, people may prefer to buy some hot tea or hot coffee. We generate two temperature sequence including the highest and lowest temperature in each day for each user.

Table 1. Different weather conditions and their labels.

Weather type	Label
Sunny	0
Light Rain	−0.5
Heavy Rain	−1
Light Snow	−1.5
Heavy Snow	−2

4 Experiment

4.1 Dataset

The dataset we used in this study is a real-world dataset [15]. It's the consumption record of consumer to use Alipay at offline stores. This dataset includes 2000 shops in different city over the country. The dataset time covers from July 1st 2015 to October 31th 2016. We selected 1057 consumers who consumed more than 120 times and more than 3 different stores.

We calculated the information entropy of user's consumption sequence according to Eq. (12).

$$H(x) = E(log_2(1/p(x_i))] = -\sum(p(x_i)log_2(1/p(x_i))), (i = 1, 2, \ldots, n) \quad (12)$$

$P(x_i)$ in Eq. (12) represent the probability of random variables event x_i. The information entropy can be used to measure the uncertainty of random variables events. The higher the information entropy of the user's consumption record sequence, the more complex and unpredictable of consumer's consumption behavior is. Then we divided the all consumer to 3 groups by the information entropy, show in Table 2. The users' consumption behavior in Group3 is most unpredictable.

Table 2. Consumers grouping according to their consumption record sequence information entropy.

	Group1	Group2	Group3
Information entropy	0∼0.5	0.5∼1.0	>1.0
Numbers of consumer	448	296	313

4.2 Baselines Comparison

In this paper, we compared the performance of the proposed method with some other baselines. These methods are:

Most Frequent (MF): We considerate the consumption frequency is a particular aspect of people's consumption behavior, so we choose the most frequent as the baseline of our data experiment.

Hidden Markov Model (HMM): HMM is a powerful statistical tool for modeling generative sequences that can be characterized by an underlying process generating an observable sequence. It's one of the most basic and extensively used statistical tools for modeling the discrete time series.

Recency [9]**:** This baseline assumes that the recently consumed items are more likely to be reconsumed.

DYRC [16]**:** This method proposes a mixed weighted scheme to recommend repeat items based on item popularity and recency effect.

Long Short-Term Memory (LSTM): LSTM introducing a memory cell and generating a unit of computation to replace traditional artificial neurons in the hidden layer of a network. With these memory cells, networks are able to overcome some difficulties with training encountered in earlier recurrent nets.

In the experiment, we set 50 hidden units in the networks, and choose MAE(Mean Absolute Error) as loss function and Adam as optimizer to train this networks. A linear function was selected as the activation function in this network. In all six methods, we set 60 as the length of training sequence. The Fig. 2 illustrates the prediction accuracy of all the baselines and MF-RNN model on the whole three groups of customers. The MF method undoubtedly got the lowest prediction accuracy, nearly to the HMM. And we can find that the neural network method has a great performance improvement over the other four methods. Finally, the model we proposed gets 83.5% prediction accuracy on the most unpredictable group and win the best perform among all six methods. The MF-RNN improve 26.0% than MF on Group3, 23.8% than HMM, 24.1% than Recency, 21.8% than DYRC, and 6.8% than LSTM.

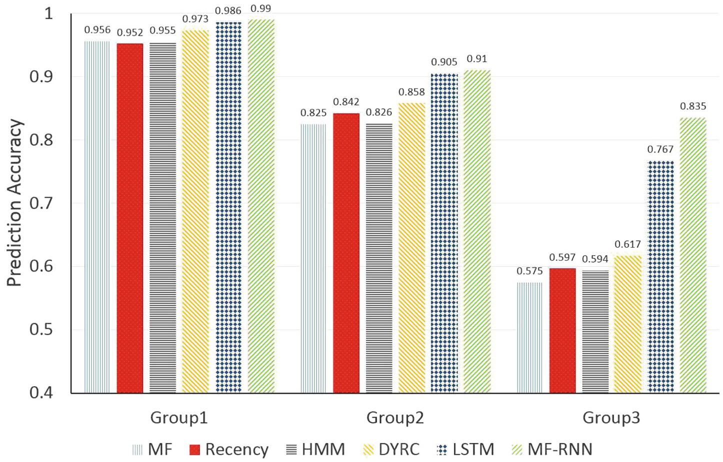

Fig. 2. Prediction accuracy comparison among six methods.

4.3 Influential Factor Analyze

In order to understand which factor has the greatest impact on consumer behavior, we compared the prediction accuracy of different influential factors on MF-RNN model. The experiment on Group3 customers shown in Table 3. We can find that the prediction model with all three factor has the best performance on the real-world dataset. The MF-RNN improve 2.5% than the RNN without any factors. This shows that the introduction of nature influential factors can improve the performance of the prediction model. The MF-RNN improve 1.8% than the non-factor RNN by introducing holiday factor, improve 1.3% by introducing weather factor, and improve 1.7% by introducing temperature factor.

Table 3. Prediction result of different influential factors on Group3.

	Non-factor (RNN)	+Holiday factor	+Weather factor	+Temperature factor	+All factor
Prediction accuracy	0.810	0.828	0.823	0.827	**0.835**

In general, consumers in different cities may have different lifestyles and lead to different daily activities. Thence, we made some data experiment to compare the repeat consumption behavior in cities of different level. We divided 313 customers in Group3 into two groups according to the city they living in. The first group includes 219 users who lives in the first and second tier cities, such as Beijing, Shanghai, Hangzhou and etc. And the second group includes 94 user who lives in other small cities. We compared the differences in repeat consumption behavior between these two groups of users. The result shows in Table 4. We can find that holiday factor has the greatest impact on users in first and second tier cities. We think this is because the pace of life in these cities is fast and most of the users in those big cities are office workers. Those users' daily consumption behavior between workdays and holidays are different. But the lifestyles in small cities are different. From the results, it can be seen that not holiday factor but the weather factor is the most important factor for the consumers in small cities.

Table 4. Prediction accuracy comparison between very large cities and small cities on Group3.

	+Holiday factor	+Weather factor	+Temperature factor
Very large cities	**0.837**	0.820	0.826
Other small cities	0.808	**0.810**	0.806

Besides, we try to analyze the consumption behavior differences in different regions. We divided the customers in Group3 into south group and north group according to their location. The south group includes 255 customers and the north group includes 58 customers. We compared the differences in consumption behavior between these two groups of users. The result shows in Table 5. It illustrates that the North China group is most sensitive to temperature factor, probably because of the extreme temperature changes in the northern China regions. And weather factor has a greater impact on South group than on north group.

Table 5. Prediction accuracy comparison between south China and north China on Group3.

	+Holiday factor	+Weather factor	+Temperature factor
South China	**0.826**	0.820	0.824
North China	0.838	0.837	**0.840**

5 Conclusion

In this paper, we proposed a prediction framework that based on MF-RNN to predict the customer's repeat consumption behavior. This method uses an three-layer RNN structure, and introduce three factors include holiday factor, weather factor and temperature factor to model customer's repeat consumption behavior. We compared the method with some other baseline methods. The experiment result shows that our MF-RNN gets better performance than MF, HMM, Recency, DYRC and LSTM. Then we compared the effect of different factors on the customers' repeat consumption behavior. And the result shows that after introduced three factors the MF-RNN get better performance, the prediction accuracy improved 2.5% than RNN without any factors. Finally, we found there is a large difference in consumption behavior between different cities and regions in China. Therefore, to a certain extent, our research has practical significance for predicting the repeat consumption behavior.

Acknowledgement. This work was supported by Zhejiang Provincial Natural Science Foundation of China (NO. LY17F020008).

References

1. Ricci, F.: Recommender Systems Handbook. Springer, New York (2011). https://doi.org/10.1007/978-1-4899-7637-6
2. Herlocker, J.L., Konstan, J.A., Borchers, A., et al.: An algorithmic framework for performing collaborative filtering. In: 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 230–237. ACM (1999)
3. Ekstrand, M.D., Riedl, J.T., Konstan, J.A.: Collaborative filtering recommender systems. Foundations and trends? Hum. Comput. Interact. **4**(2), 81–173 (2011)
4. Yi, Z., Wang, D., Hu, K., et al.: Purchase behavior prediction in M-commerce with an optimized sampling methods. In: IEEE International Conference on Data Mining Workshop, pp. 1085–1092. IEEE Computer Society (2015)
5. Russell, C.A., Levy, S.J.: The temporal and focal dynamics of volitional reconsumption: a phenomenological investigation of repeated hedonic experiences. J. Consum. Res. **39**(2), 341–359 (2011)
6. Adar, E., Teevan, J., Dumais, S.T.: Large scale analysis of web revisitation patterns. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 1197–2008. ACM (2008)
7. Adar, E., Teevan, J., Dumais, S.T.: Resonance on the web: web dynamics and revisitation patterns. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 1381–1390. ACM (2009)
8. Teevan, J., Adar, E., Jones, R., et al.: Information re-retrieval: repeat queries in Yahoo's logs. In 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 151–158. ACM (2007)
9. Anderson, A., Kumar, R., Tomkins, A., et al.: The dynamics of repeat consumption. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 419–430. International World Wide Conference Committee (2014)
10. Chen, J., Wang, C., Wang, J.: Recommendation for repeat consumption from user implicit feedback. IEEE Trans. Knowl. Data Eng. **28**(11), 3083–3097 (2016)
11. Rafailidis, D., Nanopoulos, A.: Repeat consumption recommendation based on users preference dynamics and side information. In 24th International Conference on World Wide Web, pp. 99–100. ACM (2015)
12. Zhang, F., Zheng, K., Yuan, N.J., et al.: A novelty-seeking based dining recommender system. In: 24th International Conference on World Wide Web, pp. 1362–1372. International World Wide Conference Committee (2015)
13. Lichtenthäler, C., Schmidt-Thieme, L.: Multinomial SVM item recommender for repeat-buying scenarios. In: Spiliopoulou, M., Schmidt-Thieme, L., Janning, R. (eds.) Data Analysis, Machine Learning and Knowledge Discovery. SCDAKO, pp. 189–197. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-01595-8_21

14. Lipton, Z.C., Berkowitz, J., Elkan, C.: A critical review of recurrent neural networks for sequence learning. Comput. Sci. (2015)
15. Tianchi big data contest. https://tianchi.aliyun.com/competition/index.htm. Accessed 2 May 2018
16. Benson, A.R., Kumar, R., Tomkins, A.: Modeling user consumption sequences. In: 25th International Conference on World Wide Web, pp. 519–529. International World Wide Conference Committee (2016)