# Valid Context Detection Based on Context Filter in Context-Aware Recommendation System

Shulin Cheng
Computer Engineering and Science Institute, Shanghai University Shanghai, China
The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing, China
chengshulin@shu.edu.cn

Bofeng Zhang*
Computer Engineering and Science Institute, Shanghai University.Shanghai, China
bfzhang@shu.edu.cn

Guobing Zou*
Computer Engineering and Science Institute, Shanghai University.Shanghai, China
gbzou@shu.edu.cn

Mingqing Huang
Computer Engineering and Science Institute, Shanghai University.Shanghai, China
mqhuang@shu.edu.cn

Ying Lv
Computer Engineering and Science Institute, Shanghai University.Shanghai, China
331544690@qq.com

## ABSTRACT
Context as a kind of quite important information plays a significant role in context-aware recommendation system (CARS). Many studies have been proved that context help promote to improve the effectiveness of recommendations. But a serious challenge has yet not been solved well, which is how to detect valid contexts for users in CARS, since different users have different sensitivity to contexts. Motivated by the observations, we proposed a method of valid context detection based on context filter. Context filter comprises two selection phases. In the first phase, context selection depends on the expert experiences, which is also called primary selection. We focus on the second selection phase named refinement selection based on one-way analysis of variance (ANOVA). By one-way ANOVA, a utility function is put forward to measure user's context sensitivity to detect valid contexts. We verified the effectiveness of detection method by the experiments on a small real film dataset.

## CCS Concepts
• **Information systems→ World wide web→ Web searching and information discovery→ Personalization**

## Keywords
Context; Valid Context Detection; Context-Aware Recommender System; One-way; ANOVA.


*DSIT2018*, July 20–22, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-65215/18/07…$15.00
https://doi.org/10.1145/3239283.3239318

## 1. INTRODUCTION
In recent years, with the change of the application environment, such as the widespread application of social media and the prevalence of social networks, more and more information could be available in recommendation system (RS), including rich context information such as time, space, user's mood and emotion, and interactions. RS, which comprises rich context information and generates personalized recommendation by them, is called context-aware recommendation system (CARS), and was first defined and proposed by Adomoviciuset al. [1]. The traditional RS lacks the rich context information or does not make full use of contexts, so it is difficult to break the bottleneck of performance.

Context [2][3] is any information used to depict the situation of an entity in CARS, which has an important influence on user interest and recommendation. Therefore, to improve the recommendation performance, the context information should be incorporated into recommendation. The existing literatures [1][4][5] show that the recommendation effectiveness could be improved much by context information. However, an important challenge is not solved well, namely, how to identify the valid contexts in CARS, since not all contexts are positive for recommendation. And not all users are equally sensitive to the contexts. For instance, some users are quite sensitive to time context, their interests changed with time; while other users are not sensitive to time, their interests are relatively stable for a long time. Hence, we propose the concept of valid context detection and the context filter based on one-way ANOVA for detecting the valid contexts in CARS.

## 2. RELATED WORKS
The word "context" was originally derived from a concept in the field of pervasive computing [6], which implies three aspects: where, with whom, and the resources available around it, such as location, surrounding companions, noise level, and network connectivity. The systems using context-computing help promote

* Corresponding authors: Bofeng Zhang and Guobing Zou

and coordinate the interactions between people and devices, computers, and others.

There are many different versions of the concept of context. In CARS, the “context” recognized by most scholars is the definition proposed by Dey in his two papers [2][3]. In this paper, we also refers to it, i.e., the context is any information used to characterize the situation of an entity. An entity represents an abstract concept in the system, which may be a person, location, or object that is considered to be related to the interaction between the user and the system, also including the user and the application system-itself.

Context computing system originally derived from pervasive computing system, while CARS first formally defined and proposed by Adomavicius etc. [1]. They put forward the context-aware-related paradigms and theories which extend the research on RS. Some contexts are often utilized to model user interests, mainly including time context [7][8][9][10] and social context[4][9]. Time is the context which most easily obtained in CARS, like user rating time, browsing time, publishing time. In early time, Ding [10] etc. proposed the CF recommendation based on time weight, which embodied the user sensitivity to time. In user document interest modelling, Cheng [7] etc. argued recent documents user browsed reflected the current interests and incorporated time into user modeling by a time decay function. With the application of social network platforms, social contexts are also leveraged in RS. Taking into account the user's interest instability and time-varying characteristics, Yin [8][9] etc. combined the user's own interest and time to propose a temporal context-aware hybrid model TCAM in the social media system. They extended the TCAM model and utilized social Information to reduce data sparsity in user modeling. Huang [4] etc. et al. leveraged the social context to construct user socialized model, which greatly improved the accuracy of predicting user interests. In the recommendation of e-commerce fields, Panniello [6] etc. regarded context as one dimension when studying the context-aware recommendation method. They also argued the context has hierarchy structure with K values, which are utilized to build K micro-user models.

# 3. CONTEXT FILTER FOR DETECTING VALID CONTEXTS

There are many contexts in CARS, but it is not sure that they are all valid for users or recommendations. For instance, for the context of weather in POI, if the target is to recommend the inside objects, it is not valid. But if recommending the outside objects, it is a type of context which should be considered. The system must not recommend an outside visit for users in rainy days. Again, for time context in the previous researches, considering the influence of time, a time decay function [10] as the weight was put forward to model user interests for every user in recommender systems. But in fact, not all users are sensitive to time context, namely, different users are sensitive to time context with varying sensitivity. Valid context would promote the recommendations; while the invalid context could be negative for recommendations. Hence, to improve the recommendations in CARS, we propose context filter to detect the valid contexts for users, i.e. filter the contexts to which users are not sensitive.

## 3.1 Framework of Context Filter

Our proposed context filter is mainly utilized to detect user context sensitivity, which comprises two context selection phases showed in Figure 1.

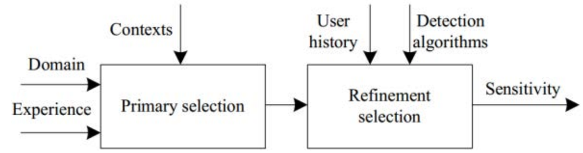


**Figure 1. Context Filter for user sensitivity on contexts.**

Since, there much context information involved in CARS and they are complex to be processed, it is hard to determine the user context sensitivity and deal with it by a unified and mature way. Therefore, in Primary selection, the initial context information is determined by expert experience. And then in Refinement selection, they will be refined by some technology methods. In other words, some methods and algorithms will be leveraged to detect the validation of each context and analyze the context sensitivity for users.

## 3.2 Formalized Utility Function of Context Filter

Refinement selection phase of context filter depends on the specific algorithm to detect whether user is sensitive to context. The key of the algorithm is the utility function f which is formalized as follows.

$$f_{utility} = f(h_{u_j}, c_i) \tag{1}$$

The function is an abstract target function which is different in line with the recommendation items, maybe based on item ratings or user interest categories. It is easy to understand if the interests of user $u_j$ are affected by context $c_i$, the function value $f(h_{u_j}, c_i)$ calculated by user history records under context $c_i$ will be also quite influenced. Therefore, we proposed a utility function based on one-way ANOVA.

## 3.3 One-way ANOVA for Building Utility Function

Assume context $c_i$ has $K_{c_i}$ context levels, $N_{c_i}^k$ denotes the rating number under the kth context level. $c_i$ is viewed as a categorical variable and can be assigned $K_{c_i}$ different values. Then $K_{c_i}$ segmentations of rating data will be extracted from the dataset by $K_{c_i}$ context levels. One-way ANOVA can be leveraged to judge the significant differences between data segmentations. The data structure of one-way ANOVA is designed as table1.

**Table1. The data structure of one-way ANOVA**

| Context level | Item rating | Average rating |
|---|---|---|
| $V_{c_i}^1$ | $R_{11}, R_{12}, \cdots, R_{1N_{c_i}^1}$ | $\bar{R}_1$ |
| $V_{c_i}^2$ | $R_{21}, R_{22}, \cdots, R_{2N_{c_i}^2}$ | $\bar{R}_2$ |
| $\cdots$ | $\cdots$ | $\cdots$ |
| $V_{c_i}^{K_{c_i}}$ | $R_{K_{c_i}1}, R_{K_{c_i}2}, \cdots, R_{K_{c_i}N_{c_i}^{K_{c_i}}}$ | $\bar{R}_{K_{c_i}}$ |

Where, $V_{c_i}^k$ denotes kth level of context $c_i$, $R_{kn}$ represents item rating, i.e., observable history data, and $k = 1,2,\cdots,K_{c_i}$ ,

$n = 1,2,\cdots,N_{c_i}$ . The average rating of kth level is $\bar{R}_k$ computed by Eq. (2)

$$\bar{R}_k = \frac{1}{N_{c_i}}\sum_{n=1}^{N_{c_i}} R_{kn} \quad (2)$$

The average rating of all the items is $\bar{R}$ as following Eq. (3).

$$\bar{R} = \frac{1}{K_{c_i}N}\sum_{k=1}^{K_{c_i}}\sum_{n=1}^{N_{c_i}^{K_{ci}}} R_{kn} \quad (3)$$

Where $N = \sum_{k=1}^{K_{c_i}} N_{c_i}^k$ . $\bar{R}$ is also equal to the average value of all average ratings at corresponding context level $V_{ci}^k$ , i.e., $V_{ci}^k = \frac{1}{K_{ci}}\sum_{k=1}^{K_{ci}}\bar{R}_k$ , $k = 1,2,\cdots,K_{ci}$ .

Let SST be total variation of item ratings, SSE be variation in groups, SSA be variation between groups. The data corresponding to one context denotes one group. SST indicates the sum of squares of deviations between observed rating data and total average rating. SSE represents the sum of squares of deviations between observed rating data and each categorical average rating corresponding to some context. And SSA denotes the sum of squares of deviations between categorical average rating and the total average rating. SST, SSE and SSA are calculated by Eq. (4)-(6).

$$SST = \sum_{k=1}^{K_{c_i}}\sum_{n=1}^{N_{c_i}^k}(R_{kn} - \bar{R})^2 \quad (4)$$

$$SSE = \sum_{k,n}(R_{kn} - \bar{R}_k)^2 \quad (5)$$

$$SSE = \sum_{k,n}(R_{kn} - \bar{R}_k)^2 \quad (5) \quad SSA = \sum_{k,n}(\bar{R}_k - \bar{R})^2 \quad (6)$$

It can be proved that there exists the following relation showed by Eq. (7) between SST, SSE and SSA.

$$SST = SSE + SSA \quad (7)$$

Faced to all the levels of context ci, if user is not sensitive to it, we can assume there is the same average value between each group. Hence, make the null hypothesis $H_0$ and build the F test statistics in Eq. (8) by SSE and SSA.

$$F = \frac{SSA/(M-1)}{SSE/(\sum_n N_{c_i}^{K_{c_i}} - M)} \quad (8)$$

F obeys the distribution with degrees of freedom $(M-1)$ and $(\sum_n N_{c_i}^{K_{c_i}} - M)$ . Under the condition of accepting null hypothesis H0, the mathematical expectations of denominator and numerator of Eq. (8) are equal to $\sigma^2$ , then the value of F is close to 1. But if reject null hypothesis $H_0$ , there exists differences between average ratings of context levels. And the expectations of numerator of Eq. (8) will be greater than $\sigma^2$ , which will become greater and greater with the incensement of sample average value of each context level. Eventually, the resulting value of F will be much greater than 1.

So, given the significant level of $\alpha = 0.05$ ,when $F > F_a(M-1,\sum N_{ci}^{K_{ci}}\text{-}M)$ , the null hypothesis $H_0$ is rejected. Then we can conclude that there exists a significant difference in the levels of context $c_i$ . That is to say the different levels of context ci have significant effect on user interest preferences.

On the basis of above analysis, statistics F based on one-way ANOVA could be viewed as the utility function $f_{Utility}$ as following Eq. (9). The judgment rule is the test rule of F hypothesis.

$$f_{Utility} = f(SSA,SSE,M,N_{ci}^{K_{ci}}) = \frac{SSA/(M-1)}{SSE/(\sum N_{ci}^{K_{ci}} - M)} \quad (9)$$

In addition, the significant correlation coefficient calculated by SSA and SST in Eq. (10) can also be used to verify the significant difference of hypothesis test.

$$r = \sqrt{\frac{SSA}{SST}} \quad (10)$$

When $0 \le r \le 1$, the greater r is, the more influence context $c_i$ has on user interest preferences, otherwise the less influence.

# 4. EXPERIMENTS

## 4.1 Dataset

Since there doesn't exist a public and standard dataset with rich contexts in CARS, in order to verify the method of valid context detection based on one-way ANOVA, we conduct the experiments on a small real dataset of LDOS-CoMoDa which is a film dataset collected by professor Odic from Ljubljana University in Slovenia. It comprises much information like users, films, ratings and rich contexts, in which the user information is in desensitization. The dataset includes 12 contexts, which are time, season, weather, social, dayType, location, endEmo, domEmo, mood, physical, decision and interaction. LDOS-CoMoDa is quite suitable for the experiments of recommendation in CARS. Obviously, the levels of all the contexts are discrete and viewed as the categorical values of context variables.

The initial statistics of LDOS-CoMoDa is showed in table 2. Most of users' ratings are few and most of items' ratings are also few. In order to better identify the user context sensitivity, we choose the users with more than 5 ratings at least for the experiments. We obtain 38 users and their corresponding ratings.

**Table2. Statistics information of LDOS-CoMoDa**

| Statistics item | Value |
|---|---|
| User num. | 95 |
| Item num. | 697 |
| Rating num. | 1101 |
| User average rating num. | 11.6 |
| Item average rating num. | 1.6 |
| Max. num. of user ratings | 144 |
| Min. num. of user rating | 1 |
| Max. num of item ratings | 22 |
| Min. num. of item ratings | 1 |

## 4.2 Experiment Design

Given a context $C_i$, if user is not sensitive to it, then there is no significant difference between the user's ratings under all the levels of context $C_i$, otherwise there is significant difference. Therefore, we build the data table of one-way ANOVA of context $C_i$, showed in table3. The different context levels are viewed as context categorical variable values. The average rating of each item type is regarded as the element of data under the context level. There are 25 item types (i.e. film genres). In reality, most of users are interested in few types of items.

Given context $C_i$, firstly, build the null hypothesis, namely, user is not sensitive to context $C_i$. That is to say, there not exists significant difference between average ratings of all item types under context $C_i$. The significant level pvalue is 0.05. Then construct the statistics F of one-way ANOVA to detect each user's sensitivity to context ci.

**Table3. The data table of one-way ANOVA for context $c_i$**

| Context $c_i$ | Average ratings of each item type | Average value of context level |
|---|---|---|
| $c_{i1}$ | $\bar{R}_{11}, \bar{R}_{12}, \cdots, \bar{R}_{1,26}$ | $\bar{R}_1$ |
| $c_{i2}$ | $\bar{R}_{21}, \bar{R}_{22}, \cdots, \bar{R}_{2,26}$ | $\bar{R}_2$ |
| $\cdots$ | $\cdots$ | $\cdots$ |
| $c_{iK}$ | $\bar{R}_{K1}, \bar{R}_{K2}, \cdots, \bar{R}_{K26}$ | $\bar{R}_K$ |

## 4.3 Experimental Results

### *4.3.1 Detection Results*

We obtained each user's context sensitivity by context filter. But we only provide the detection results of one in user set as a representative, in which the user has the maximum rating number, since it is not appropriate to exhibit every user's detection data. The details are showed in table4.

**Table 4. The detection results of the user with maximum rating number**

| Context | SST | SSE | SSA | F | r | Sensitivity |
|---|---|---|---|---|---|---|
| time | 128.24 | 109.36 | 18.88 | 1.8420 | 0.3837 | N |
| daytype | 83.84 | 68.43 | 15.41 | 2.7018 | 0.4287 | N |
| season | 110.80 | 82.07 | 28.73 | 3.7336 | 0.5092 | **Y** |
| location | 77.19 | 44.75 | 32.43 | 11.5959 | 0.6482 | **Y** |
| weather | 130.64 | 113.48 | 17.17 | 1.6135 | 0.3625 | N |
| social | 102.47 | 48.68 | 53.79 | 13.2584 | 0.7245 | **Y** |
| endEmo | 223.90 | 187.54 | 36.36 | 1.8611 | 0.403 | N |
| dominantEmo | 253.49 | 212.70 | 40.79 | 1.7898 | 0.4011 | N |
| mood | 0.00 | 0.00 | 0.00 | 0.0000 | 0.0000 | N |
| physical | 59.80 | 38.20 | 21.60 | 9.0466 | 0.6010 | **Y** |
| decision | 60.50 | 58.50 | 2.00 | 0.5462 | 0.1817 | N |
| interaction | 49.05 | 47.94 | 1.10 | 0.368 | 0.1499 | N |

Where F denotes F test statistic, r represents the correlation coefficient of test statistics. The greater the value of F, the more sensitive to context $C_i$ the user is. Observe from table4. The data in table 4 shows that the user with the maximum rating number are sensitive to 4 of 12 contexts, which are season, location, social and physical. The rows of 0 values in table4 denote the corresponding users rate items only in one level of context c. For example, in table 4 for context mood, the user with maximum ratings rates films under the context level of "Positive", and not rates any films under the other two context levels of "Neutral" and "Negative". Naturally, the detection of context mood is not sensitive. Therefore, we can find out 4 of 12 contexts are valid for the user with maximum ratings, the others are not.

### *4.3.2 User Context Sensitivity on Total*

The overall sensitivities of users to 12 contexts in LDOS-CoMoDa dataset are demonstrated in Figure 2. The contexts with the most sensitivity is mood, social and weather, accounting for 52.63% of the total users, respectively. The second sensitive contexts are location, endEmo and time, accounting for 47.37%, 47.37% and 44.74% of total users. The rest 6 contexts are sensitive for users in some extent.

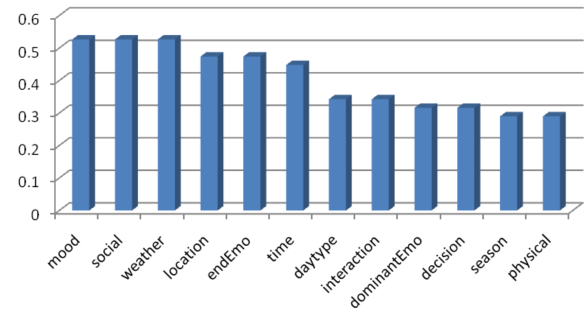


**Figure 2. The overall sensitivities of users.**

The results of Figure 2 are interpretative, which are consistent with reality. Take the social context for example, it comprises 7 context levels of Alone, My partner, Friends, Colleagues, Parents, Public, and My family. When user watches films, the film types user selected are very different with varying companies. With companies, users commonly consider their preferences. Hence, for the user, the watching films are distinguishing with those when alone, maybe they are not his most interested, just due to considering the feelings of his companies. The situations result in the fluctuations of user ratings which embody the sensitivity to the context of "social". The other sensitive contexts are similar interpretative, the details are not repeated.

### *4.3.3 Statistics of User Sensitivity to Multiple Contexts*

Different users are sensitive to different context with varying sensitivities. Even if for the same user, he/she is also differently sensitive to various contexts, and maybe simultaneously sensitive to multiple contexts with different sensitivities, which are showed in Figure 3. The abscissa represents the context number to which users are simultaneously sensitive, and the ordinate denotes the user number. It shows that 9 users who are strongly sensitive to contexts are simultaneously sensitive to 8 contexts; both 3 users are, respectively, simultaneously sensitive to 4 contexts and 6 contexts; only 6 users are sensitive to one context. Therefore, there exists quite differences between users who are simultaneously sensitive to multiple contexts. It also demonstrates the context sensitivity differences of users. In others words, the context validations are distinguishing to different users. Hence, we can conclude that there exists greater difference between numbers of contexts to which users are simultaneously sensitive.

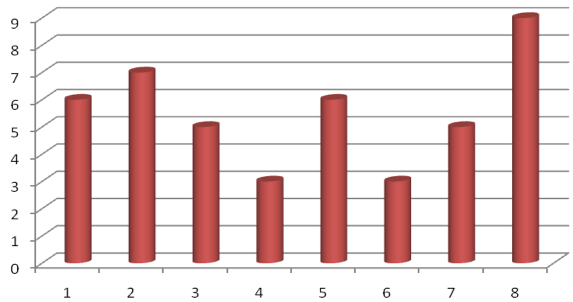


**Figure 3. Statistics about users with simultaneous sensitivity to multiple contexts.**

# 5. CONCLUSIONS AND FUTURE WORK

In this work, we focus on the valid context detection in CARS, which, to our best of knowledge, is the concept first put forward by us. We proposed a detection method based on the context filter which includes primary selection and refinement selection of contexts. Primary selection depends on expert experiences in some specific domain. We discussed the refinement selection of context in detail, in which a utility function based on one-way ANOVA is put forward to detect user valid contexts. By experiments, we verified different users have different sensitivity on contexts. Even if for the same user, he/she has varying sensitivity to different contexts. Therefore, it is indispensable to detect the valid contexts for users before generating recommendations. The method we present in this paper is by no means finished, but rather work-in-progress. In the future, we try the method of ANOVA based on multiple factors rather than one by one context. And we intend to conduct more experiments on different types of datasets with rich context information.

# 6. ACKNOWLEDGEMENTS

We would like to thank all of the anonymous reviewers for their insightful comments. This work was partially sponsored the National Key Research and Development Program of China (Grant No. 2017YFC0907505), the National Natural Science Foundation of China (Grant No. 61772128), the Shanghai Natural Science Foundation (Grant No. 18ZR1414400 and 17ZR1400200), the Natural Science Research Foundation of the Education Department of Anhui Province of China(Grant No. KJ2018A0382), and the Outstanding Young Talents Program in Colleges and Universities of Anhui Province (Grant No. gxyqZD2018060) of China.

# 7. REFERENCES

[1] Adomavicius, Gediminas, and Alexander Tuzhilin. 2005. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." IEEE transactions on knowledge and data engineering 17(6): 734-749.

[2] Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. 1999. Towards a better understanding of context and context-awareness. In International Symposium on Handheld and Ubiquitous Computing (pp. 304-307). Springer, Berlin, Heidelberg.

[3] Dey, A. K. 2001. Understanding and using context. Personal and ubiquitous computing, 5(1), 4-7.

[4] Huang, M., Zhang, B., Zou, G., Gu, C., Wu, P., Cheng, S., & Zhang, Z. 2016. Socialized User Modeling in Microblogging Scenarios for Interest Prediction. In Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 2016 IEEE 14th Intl C ,pp. 124-131.

[5] Deng, S., Wang, D., Li, X., & Xu, G. 2015. Exploring user emotion in microblogs for music recommendation. Expert Systems with Applications, 42(23), 9284-9293.

[6] Panniello, U., & Gorgoglione, M. 2011. A contextual modeling approach to context-aware recommender systems. In Proceedings of the 3rd Workshop on Context-Aware Recommender Systems.

[7] Cheng, S., & Liu, Y. 2015. Time-Aware and Grey Incidence Theory Based User Interest Modeling for Document Recommendation. Cybernetics and Information Technologies, 15(2), 36-52.

[8] Yin, H., Cui, B., Chen, L., Hu, Z., & Huang, Z. 2014. A temporal context-aware model for user behavior modeling in social media systems. In Proceedings of the 2014 ACM SIGMOD international conference on Management of data, pp. 1543-1554.

[9] Yin, H., Cui, B., Chen, L., Hu, Z., & Zhou, X. 2015. Dynamic user modeling in social media systems. ACM Transactions on Information Systems (TOIS), 33(3), 10.

[10] Ding, Y., & Li, X. 2005. Time weight collaborative filtering. In Proceedings of the 14th ACM international conference on Information and knowledge management, pp. 485-492.