Advances in Computer Science Research (ACSR), volume 61
7th International Conference on Education, Management, Computer and Society (EMCS 2017)

Research on Information Fusion Method of Mobile Commerce based on Monte Carlo Algorithm under Big Data Environment

Tingbin Chen[1, a], Qisong Zhang[1, b] and Jingshu Wang[1, c,*]

[1] Dalian Neusoft University of Information, Department of Information Management, Dalian, China

[a]chentingbin@neusoft.edu.cn, [b]zhangqisong@neusoft.edu.cn, [c]wangjingahu@neusoft.edu.cn

*The corresponding author

Keywords: Big data; Location information; Information fusion; Improved monte carlo algorithm

Abstract. In the large data environment, the amount of mobile e-commerce information increases sharply and highlights the characteristics of users' location. How to utilize the user's location quickly and effectively to find the valuable information of the users and supply them with the decision-making service becomes one of the important subjects of the mobile electronic commerce in recent years. Based on the traditional Monte Carlo algorithm, this paper proposes an improved Monte Carlo algorithm, a fast Monte Carlo algorithm, which can effectively solve the problem of particle degradation in the traditional algorithm. Finally, the simulation experiment is carried out. Compared with the traditional information fusion method, the experiment demonstrates that the fusion method has high accuracy and reduces the complexity of the system algorithm, so that the e-commerce platform can predict the users' behavior information more accurately and can provide decision-making services more appropriately.

Introduction

With the increasing amount of mobile e-commerce data and the diversification of data structure, there comes the era of large data. How to find the effective data quickly and effectively in massive, heterogeneous or non-heterogeneous data which directly or indirectly influences the needs of users plays an increasingly important role. In particular, in recent years, based on the combination of location-based services and mobile e-commerce, to provide the decision- making for the users based on users' location becomes one of the important trends of the development of mobile e-commerce in the future. [1] How to adapt to the users' mufti-source location [2] and provide users with decision-making has become the most conspicuous technical points. In this paper, a mufti-source information fusion method based on improved Monte Carlo algorithm which is proposed to improve the value of mufti-user location information and promote the development of mobile e-commerce push service in the large data environment has great practical significance.

Traditional Monte Carlo Algorithm

Monte Carlo algorithm, also known as statistical simulation method, random sampling technology, is a stochastic simulation method, based on probability and statistical theory, uses random number (or pseudo-random number) to solve a lot of problems of calculation [3]. When the solution is the probability of a random event, or the expected value of a random variable, the probability of this random event is estimated by the frequency of the occurrence of such an event by some "experimental" approach, or getting some of the numerical features of the random which are treated as a solution to the problem.

The essence of the Monte Carlo algorithm is the deformation of the Bayesian probability algorithm, which is based on a set of discrete particles with weights to simulate the posterior probability density of the estimated system state and to predict the state of the system, the system provides the corresponding decision-making support. In this paper, the users' location information in mobile e-commerce is sampled and the user's location information which is improved by Monte Carlo algorithm is fused so that it can intelligently provide the decision-making support for the e-commerce users [4,5].

Previous work shows that acts as one of the last technologies, big data plays an important role in fields like in e-business, e-finance and supply chain. However, rare scholars propose models for catching, dealing and presenting supply chain data which is volume, structured and unstructured.

Bayesian Filter Theory. Firstly, it is assumed that the state of the dynamic system estimated by X in the Bayesian filter theory which is based on the users' position data obtained from the LBS system. Specifically, in the mobile e-commerce, the dynamic system contains the mobile users' behavior and their geographical information, the state refers to the user's decision-making estimate and the observation data mainly contains the geographical location information from the LBS system and users' real-time behavior information. The Bayesian filtering theory assumes that the system state conforms to the Markovian rule, that is, if the current system state is known, then the system state at the next moment is related only to the current state, regardless of the information at the previous time. The central idea of Bayesian filtering is to estimate the probability density of the state space based on the test data. In this paper, the posterior distribution of the user state is called the credibility, and is expressed by the formula (1):

$$Bel(x_t) = p(x_t / z_{0\cdots t}) \tag{1}$$

Where x is the system state and x_t is the system state at t time. Where the $z_{0\cdots t}$ system state is the system state of the time, the behavior information data and the user location information data obtained from the LBS system at time 0 to time t, the former can be represented by o, which can be represented by a, so (1) can be expressed as follows:

$$Bel(x_t) = p(x_t / o_t, a_{t-1}, o_{t-1}, a_{t-2}, \cdots, o_0) \tag{2}$$

In order to lose the generality, it is assumed that the position data and the behavior data are not arriving at the same time, the behavior information arrives first, the position information is subsequently arrived. o_t represents the behavior information of the mobile user at time t, and a_{t-1} represents the location information of the mobile user at time t-1.

The initial confidence obtained by the Bayesian filtering recursion can describe the initial state of the dynamic system state: when the initial state of the dynamic system is unknown, it is described by a uniform distribution; when the initial state of the dynamic system is known, Described by the Gaussian distribution. Therefore, equation (2) can be transformed into equation (3):

$$Bel(x_t) = \frac{p(o_t / x_t, a_{t-1}, \cdots, o_0) p(a_{t-1}, \cdots, o_0)}{p(o_t / a_{t-1}, \cdots, o_0)}) \tag{3}$$

It is not difficult to see from the above formula that the denominator of formula (3) does not change relative to the variable and can therefore be defined as follows:

$$Bel(x_t) = \eta p(o_t / x_t, a_{t-1}, \cdots, o_0) p(a_{t-1}, \cdots, o_0) \tag{4}$$

Where η is a constant, as shown in equation (5):

$$\eta = p(o_t / a_{t-1}, \cdots, o_0)^{-1} \tag{5}$$

It follows that the current state of the dynamic system is only relevant to the previous moment and is consistent with the Marcoff hypothesis. Marcoff can be expressed as:

$$p(o_t / x_t, a_{t-1}, \cdots, o_0) = p(o_t / x_t) \tag{6}$$

Equation (6) can be changed as follows:

$$Bel(x_t) = \eta p(o_t / x_t) p(x_t / a_{t-1}, \cdots, o_0) \tag{7}$$

(7) to the right of the integral, obtained (8):

$$Bel(x_t) = \eta p(o_t / x_t) \int p(x_t / x_{t-1}, a_{t-1}, \cdots, o_0) p(x_{t-1} / a_{t-1}, \cdots, o_0) dx_{t-1} \tag{8}$$

According to the Markov hypothesis, it can be simplified as follows:

$$p(x_t / x_{t-1}, a_{t-1}, \cdots, o_0) = p(x_t / x_{t-1}, a_{t-1}) \tag{9}$$

Thus equation (7) can be expressed as follows:

$$Bel(x_t) = \eta p(o_t / x_t) \int p(x_t / x_{t-1}, a_{t-1}) p(x_{t-1} / a_{t-1},, o_0) dx_{t-1} \tag{10}$$

According to the definition of the credibility of the formula, from (10) to obtain (11):

$$Bel(x_t) = \eta p(o_t / x_t) \int p(x_t / x_{t-1}, a_{t-1}) Bel(x_{t-1}) dx_{t-1} \tag{11}$$

Equation (11) for the recursive update formula can be used to recursive state estimation of the system, as shown in Fig. 1.

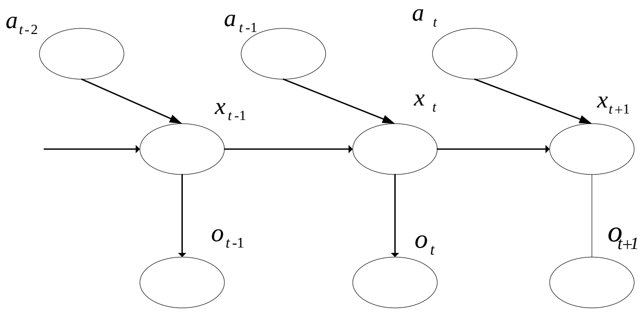

Figure 1. Schematic diagram of the dynamic system

From Figure 1 can be drawn:

(1) When t = 0:

The update density of the dynamic system is:

$$p(x_0 / o_0) = \frac{p(o_0 / x_0) p(x_0)}{p(o_0)} \tag{12}$$

The predicted density of the dynamic system is:

$$p(x_1 / o_0) = \int p(x_1 / x_0) p(x_0 / o_0) dx_0 \tag{13}$$

(2) When t> 0:

The update density of the dynamic system is:

$$p(x_t / o_{1:t}) = \frac{p(o_t / x_t) p(x_t / o_{1:t-1})}{p(o_t / o_{1:t-1})} \tag{14}$$

The predicted density of the dynamic system is:

$$p(x_{k+1} / o_{1:k}) = \int p(x_{t+1} / x_t) p(x_t / o_{1:t}) dx_t \tag{15}$$

Monte Carlo Principles. The core idea of Monte Carlo algorithm is to observe m samples and then describe the credibility of the series of samples $Bel(x)$:

$$Bel(x) \approx \{x^{(i)}, w^{(i)}\}_{i=1,2,\cdots m} \tag{16}$$

Where $x^{(i)}$ indicates the i-th sample of random variable x, $w^{(i)}$ is the weight of the i-th sample, which is the degree of importance, the discrete probability density of credibility $Bel(x)$ can be approximated by a series of samples.

The initial state of the dynamic system is represented by $Bel(x_0)$ and it can be described by the initial set of observation samples. In this paper, the user's initial location is unknown in mobile e-commerce, it may be in a particular space (a particular city) in any location, according to 1.1 we can know that when the initial state of the dynamic system is unknown, the system state can ben described of uniform distribution description and the important parameter of each behavior is $1/m$.

The system recursively updates as follows:

(1) extract the sample from the credibility $Bel(x_{t-1})$ of the sample set at time t-1, whose probability density is subordinate to $Bel(x_{t-1})$.

(2) Generate sample $x_t^{(i)}$, subject to the expression of $p(x_t / x_{t-1}^{(i)}, a^{t-1})$, $< x_t^{(i)}, x_{t-1}^{(i)} >$, based on::

$$q_t = p(x_t / x_{t-1}, a_{t-1}) * Bel(x_{t-1}) \tag{17}$$

Where q_t is proposed distribution which means making recommendations of proposed posterior distribution of samples

(3) Revise q_t and re-express sample $< x_t^{(i)}, x_{t-1}^{(i)} >$:

$$Bel(x_t) = \eta p(o_t / x_t^{(i)}) p(x_t^{(i)} / x_{t-1}^{(i)}, a_{t-1}) Bel(x_{t-1}^{(i)}) \tag{18}$$

Where

$$w^{(i)} = p(o_t / x_t^{(i)}) \tag{19}$$

The quotient of distribution of the target distribution (18) divided the proposed distribution (17) is proportional to the important factor $w^{(i)}$, η Is the scale factor

$$\frac{\eta p(o_t / x_t^{(i)}) p(x_t^{(i)} / x_{t-1}^{(i)}, a_{t-1}) Bel(x_{t-1}^{(i)})}{p(x_t / x_{t-1}, a_{t-1}) * Bel(x_{t-1})} = \eta p(o_t / x_t^{(i)}) \tag{20}$$

Repeat sampling m times it can be $x_t^{(i)} (i = 1, \cdots, m)$。Then the important factor $w^{(i)}$ is normalized so that their sum is 1, thus defining the discrete probability distribution.

The Shortcomings of the Traditional Monte Carlo Algorithm. One of the shortcomings of the traditional Monte Carlo algorithm is the phenomenon of particle degradation, that is, after several iterations of the dynamic system, most of the sampled particles in the system have changed the weight, and the weight of most sampled particles is very small, but the number of sampled particles The weight will become larger. In this paper, the variance of the particle weight becomes larger, which makes the smaller weight of the system lose its effect on the system. In order to judge the weight of the sampled particle reasonably, this paper adopts the formula (21) to judge effectively:

$$N_{eff} = \frac{N}{1 + Var(w_k^i)} \tag{21}$$

$w_k^i = p(x_k^i / z_{1:k}) / q(x_k^i / x_{k-1}^i, z_k)$ represents the true weight of the sampled particles.In the actual operation, it is difficult to find out N_{eff}, but the study found that an approximate expression can be used to estimate N_{eff}, the approximate expression is:

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} w_k^i} \tag{22}$$

N_{eff} Represents the degree of particle degradation, The smaller the value of N_{eff}, the smaller the degree of particle degradation, and vice versa.From the equation (21) is not difficult to see, when the number of particles N is large, the N_{eff} corresponding will be great, but in the actual operation, it is difficult to happen.

It can be seen from equation (21) that in order to make the increase of N_{eff}, only to achieve the minimum of $Var(w_k^i)$, it is essential to select the importance of density distribution $q(x_k / x_{k-1}^i, z_k)$, x_{k-1}^i and z_k is known, the optimal importance of density distribution:

$$q(x_k / x_{k-1}^i, z_k)_{opt} = p(x_k / x_{k-1}^i, z_k) = \frac{p(z_k / x_k, x_{k-1}^i) p(x_k / x_{k-1}^i)}{p(z_k / x_{k-1}^i)} \tag{23}$$

$$w_k^i = w_{k-1}^i p(z_k / x_{k-1}^i) \tag{24}$$

If x_{k-1}^i is known, the value of w_k^i is equal for the system sample, $Var(w_k^i) = 0$, Which leads to the importance of density function for the optimal density function. But in practice, it is difficult to sample $p(x_k / x_{k-1}^i, z_k)$, the current mainstream research method is to make, $p(x_k / x_{k-1}^i, z_k) = p(x_k / x_{k-1}^i)$.However, this method is easy to lose the location information of the user at time K, so that the location information in the mobile e-commerce user information fusion process does not play a role, and ultimately may lead to the next step can not predict the user behavior, resulting in the user can not push effective Behavior information.Therefore, the traditional Monte Carlo algorithm is improved to make it suitable for mobile e-commerce.

The Improved Monte Carlo Algorithm

Fast Monte Carlo algorithm. It is difficult to carry out the real operation because of the problem of particle degradation in the traditional Monte Carlo algorithm from 1.3 . In view of the defects of the traditional Monte Carlo algorithm, the researchers have made a lot of research work to improve the work, Emphasis on the re-sampling method to improve, but this improvement does not reduce the complexity of the algorithm to make the system more optimized, and if blindly reduce the complexity of the algorithm may lead to the accuracy of information fusion, so find a guarantee Mobile e-commerce system platform information fusion accuracy under the premise of reducing the complexity of the algorithm to become e-commerce platform system research focus.In this paper, we propose an improved Monte Carlo algorithm, the fast Monte Carlo algorithm, which first uses the unscented Kalman filter to generate the importance of the proposed distribution in the sample stage, so that the sampling distribution is similar to the true distribution, and the particles in the traditional algorithm The problem of degradation.

Secondly, the adaptive algorithm is adopted, and the steps of resampling are omitted to ensure the diversity of particles in the system. Finally, in order to prevent the occurrence of particle depletion occurs, using the Markov chain Monte Carlo algorithm in the sampling MH algorithm. Based on the traditional Monte Carlo algorithm, this algorithm combines a variety of intelligent algorithms and sampling methods to ensure that the complexity of the system algorithm is reduced without affecting the accuracy of information fusion, so that the e-commerce platform can more accurately predict User behavior information to better provide decision support services for users.

Fast Monte Carlo algorithm specific steps are as follows:

(1) Initialization: according to the traditional Monte Carlo algorithm to initialize the system;

(2) Sampling stage: at each moment for the sample of the first sample point $x_{k-1}{}^{(i)}$ $x_k{}^{(i)}$ $p_k{}^{(i)}$ with unlicensed Kalman filter update , get the proposed distribution function is:

$$q(x_k{}^{(i)} / x_{k-1}{}^{(i)}, z_{1:k}) = N(x_k{}^{(-i)}, p_k{}^{(i)}) \tag{25}$$

(3) Calculate the weight: use the traditional Monte Carlo algorithm to calculate the particle weight in the system;

(4) Resampling: The sample is sampled by the method $MCMC$. According to the Markov Monte Carlo principle, each particle in the collection according to the weight is MH sampled again, and the importance weight of each particle is assigned to $1/N$, and then repeat the (2), (3), (4) until the end of fusion.

The Information Fusion Process Based on Improved Monte-Carlo Method. In this thesis, improved Monte-Carlo Method can predict purchasing behavior's reliability $Bel(x_k)$ in mobile e-commerce, which used N random samples with weights. The sample space is $S=\{s_i / i=1,2\cdots N\}$. Discretizing the sample set can approximately predict the reliability of user purchasing behavior. Each sample includes a $X_i=(x_i, y_i, \theta_i)$ at the sample space inside, representing the purchasing behavior and the probability of purchasing behavior in the future, where $\sum w_i=1$. Improved Monte-Carlo Method includes purchasing behavior updating and location updating.

(1) purchasing behavior updating

Purchasing behavior updating is a convolution of historical purchasing behavior and location information distribution of mobile users. At the sample space S_k, the sample $s^i{}_k\,(i=1,2,\cdots N)$ is selected one by one. According to historical purchasing behavior $p(l / a, l')$, a new sample s' has been random generated. Then, s' is replaced by $s^i{}_k$.There were systematic error and unsystematic error. Each sample can't generate the same purchasing behavior. After the transformation, a new sample space S_{k+1} is generated, which represents the reliability of purchasing behavior at the next moment.

(2) location updating

Using LBS system, E-commerce gets a number of new sample locations. Then, the reliability of user's position from the GIS system multiply by existing behavior's reliability, the process can be divided into two steps:

a.Update the samples weights. We use the probability of each sample location in order to update all samples according to the coordinates from the GIS system.

b.Normalization of new sample weights. Replacing the old sample set by sampling, we get a new sample set. In the old sample space, the greater sample weight can improve the probability of extracting. Because the new sample is based on sample weight, the probability of new sample is proportional to the old sample weight. Obviously, new samples based on the above methods need to be normalized. The code of improved Monte-Carlo Method display as follows:

Prediction update:

For each i=1,…,N_k

Sampling $\overline{X}_k{}^{(i)}$ from probabilistic model $p(X_{k-1}{}^{(i)}, a_{k-1})$

$$\overline{w}_k{}^{(i)}=1 / N_k$$

$$S_k'=\{<\overline{X}_k{}^{(i)}, \overline{w}_k{}^{(i)}> / i=1,\cdots,N_k\}$$

Observation update：

For each i=1,…,N_k

$$\tilde{X}_k{}^{(i)}=\overline{X}_k{}^{(i)}$$

$$\hat{w}_k{}^{(i)}=\overline{w}_k{}^{(i)} * p(Z_k / \overline{X}_k{}^{(i)})$$

$$\tilde{w}_k{}^{(i)}=\frac{\hat{w}_k{}^{(i)}}{\sum_{j=1}^{N_k}\hat{w}_k{}^{(j)}}$$

$$S_k''=\{<\tilde{X}_k{}^{(i)}, \tilde{w}_k{}^{(i)}> / i=1,\cdots,N_k\}$$

Importance Resampling：

For each i=1,…,N_k

Importance Resampling from S_k'' :

$$p(X_k{}^{(j)}=\tilde{X}_k{}^{(i)})=\tilde{w}_k{}^{(i)})$$

$$X_k{}^{(i)}=\tilde{X}_k{}^{(i)}$$

$$w_k{}^{(i)}=1 / N_k$$

$$S_k=\{<X_k{}^{(i)}, w_k{}^{(i)}> / i=1,\cdots,N_k\}$$

Simulation Experiment

In this thesis, The Monte-Carlo Method, including MCL_UKF and MCL_MCMC, is compared with the development method. The simulation experiment was carried out 50 times, Simulation time is set to T=60, particle number is 100. Simulation results display in fig.2. The dotted line is real purchasing behavior and solid line is prediction of purchasing behavior. The results are shown in the figure below.

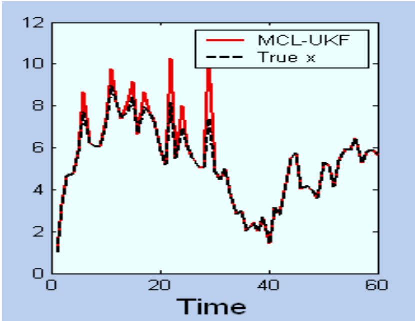

Figure.2 (a) Comparison between MCL-UKF and user behavior

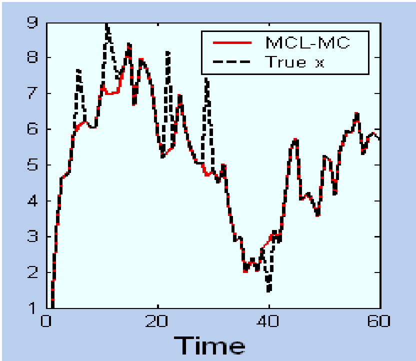

Figure.2 (b) Comparison between MCL-MC and user behavior

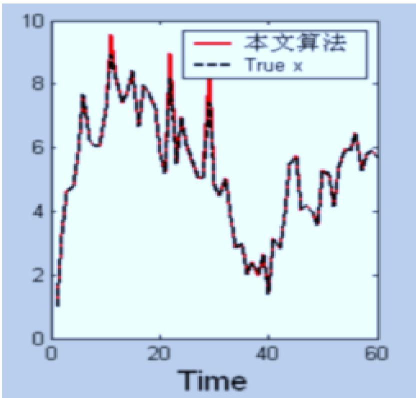

Figure 2. (c) Comparison between proposed algorithm and user behavior

In Fig. 2, Abscissa represents simulation time; Ordinate represents user behavior type in the e-commerce platform, including diet, accommodation, leisure entertainment, living service, tourism and shopping, etc. The mean and variance of three methods are given in Table 1.

Table 1 results comparison

Method	Mean	Variance	Single time
MCL-UKF	0.28835	0.007223	3.8773
MCL-MC	0.48456	0.036754	1.8990
Proposed algorithm	0.07634	0.004987	1.5890

From the Table1, we can see that the variance of this algorithm is far below the others, and the single time has the same rules. This algorithm reduces the amount of computation under the same accuracy, which has proved the algorithm is effectiveness and efficiency.

Conclusion

Firstly, this paper introduces the Bayesian filter theory, which is the basis of Monte-Carlo Method. It can be found that Monte-Carlo Method is the deformation of Bayesian filter theory. Then, the traditional Monte-Carlo Method is introduced and we point out that the algorithms exist the particle degeneration. To rectify this problem, this paper puts forward the improved Monte-Carlo Method. During the stage of sampling, this algorithm use Unscented Kalman filter to build Importance of the proposed distribution,which makes the distribution of the sampling is similar to the real. This process can be resolved the problem of particle degradation. Then, adaptive algorithm is used to process data, which resampling is avoided. And use MH sampling algorithm to avoid the particle degeneracy phenomenon. Finally, this paper uses simulation experiment to compare this algorithm with others. The results show that this algorithm reduces the amount of computation under the same accuracy. It is proved that the algorithm is effectiveness and efficiency.

References

[1] ZHANG Guangqian, LEI Caihua, LV Xiaomin. The Survey of E-Commerce Recommendation [J].Journal Of Intelligence, 2011, 30(12):60-65

[2] PAN Quan, WANG Zengfu. Basic methods and progress of information fusion (Ⅱ) [J]. Control Theory and Applications, 2012, 29(10):45-67

[3] Wang Jie, Wang Hongyu, Gao Qinghua． Enhanced Monte Carlo Localization and Tracking Algorithm for Mobile Wireless Sensor Network [J]. Journal of Electronics and Information Technology, 2010, 32(4):864-868

[4] WANG Yan，ZHANG Qiuxia. Research on Recommendation System Based on Ontology in Cloud Computing Platform [J]. Computer Development and Applications, 2013, 26(10):36-39

[5] WANG Haiyan, ZHOUYang. Qulity of Recommendation Based Trust-aware Recommender System[J],Computer Science,2014,31(5):37-41

[6] Liu Fengxia, Sun Jiarong. Designing E-Commerce Recommendation System Based On Merchandise Category [J], Computer Applications and Software, 2014, 31(5):37-41

[7] MA Hua,HU Zhigang, ZHANG Hongyu． Personalized Recommendation Framework for Trustworthy Services in Cloud Paradigm [J], Journal of Chinese Computer Systems, 2014, 32(05):12-32

[8] LIU Jing, LI Yan． The Multi-source Associated Personalized Recommendation Architecture of Mobile Ecommerce [J], Information Studies: Theory and Application, 2014, 27(04):31-43