International Journal of Hybrid Information Technology
Vol.8, No.3 (2015), pp.55-62
http://dx.doi.org/10.14257/ijhit.2015.8.3.06

Points of Interest Recommendation Based on Context-aware

Fung Wong[1], Shijun Lee[1*] and Quanrui Wong[2]

[1.] Computer School, Wuhan University, Wuhan, 430072, China
[2.] School of Information and Engineering, Henan Institute of Science and Technology, Xinxiang, 453003, China
[*]Shijun Lee, afengcom@163.com

Abstract

Existing points of interest recommendation systems do not consider either spatial-temporal factors etc or the historic behavior of users. Though both of the two kinds of models have their strength, they don't make evaluation for accuracy rate of POIs recommendation, especially recommendation Satisfaction Index (RSI) of users. In this paper, a POIs recommendation model based on context-aware is built by combining the two ideas and it makes evaluation for RSI. Experimental results showed that, compared with the existing models, the performance of the model has large improvement on the aspects of accuracy rate and RSI.

Keywords: *context-aware, POIs recommendation, recommendation system*

1. Introduction

It is a strong support for recommending Points of Interests (POIs) to users by extracting context elements from micro-blog (*e.g.,* twitter or Sina micro-blog) which could mine and analyze behavioral traits of users.

In recent years, the research of location-based recommendation system is an extension to traditional LDA model. LCARS model [1] (a recommendation system based on Location context-awareness) and STT model [2] (a recommendation system based on spatial-temporal topic modeling) are the two most representative and latest comparison models.

Recently, location-based modeling has two approaches: 1.By analyzing the history of user activity, mining the user's interests and preferences, a region model based on location awareness is built; 2.By extracting spatial-temporal elements and user profiles from check-in time, a topic model based on location context-awareness is built. The two models have advantages as well as disadvantages: LCARS just only contains location awareness information, but it overlooks the effects of spatial-temporal elements on POIs recommendation. However, STT just only contains spatial-temporal topic modeling, but it overlooks the effects of user's historic behavior on POIs recommendation.

Therefore, in this paper the two kinds of thoughts are combined for POIs Recommendation based on Context-aware so that the accuracy rate of POIs Recommendation could be improved. Points of Interests (POIs) are geographical location points that users are interested in.

2 Context-awareness Model

Context-awareness is a novel technology that the elements of time, space, user's history behavior could be extracted and sorted by some spatial-temporal series rules. Then these elements could be analyzed and mined to obtained user's personalized features for personalized recommendation, *e.g.,* POIs recommendation.

ISSN: 1738-9968 IJHIT

2.1 Awareness of Time, Space and User Behavior

Time awareness is a new method that the check-in time from micro-blog could be extracted by time series rules;

Space awareness is a recent approach that GPS coordinates shared by users from micro-blog could be sorted by spatial series rules and organized into areas and trajectories of user activity;

User behavior awareness is a novel technology that both the check-in time and shared location are considered for context-awareness association mining referring to spatial-temporal series and causal association rules.

2.2 The Effects of Region on Space Awareness

Different regions have different influence on space awareness. Some regions have greater influence for more frequent human activities and denser POIs distribution (*e.g.*, bustling commercial heart). However, some regions have lesser influence for sparse population density and POIs distribution, and weaker regional social function (*e.g.*, residential district). Thus, a hypothesis is proposed that if a region is marked more POIs and its area is tiny, then space awareness intension of this region is greater; otherwise, the intension is inferior. Based on this hypothesis, the distribution of region vs. space awareness obeys a standardized Gaussian distribution. So (1) is presented for computing regional division in given GPS coordinates as follows,

$$p(R \mid L) = \sqrt{\frac{\sum_{l=0}^{L} \exp(-\frac{1}{2}(r^2_{l,0}+r^2_{l,1})) - \sum_{l'=0}^{L} \exp(-\frac{1}{2}(r^2_{l',0}+r^2_{l',1}))}{\sum_{l=0}^{L}\sum_{l'=0}^{L}\sqrt{\pi((r_{l,0}-r_{l',0})^2+(r_{l,1}-r_{l',1})^2)}}} \quad (1)$$

where, $r_{l,0}$, $r_{l,1}$, $r_{l',0}$, $r_{l',1}$ are respectively the longitude and latitude of point l and l' in region R, L is a set of GPS coordinates in R, and p(R|L) is the distribution probability of these GPS coordinates in R computed by (1).

2.3 Formalized Description of Context-awareness Model

Generative process of the model is described below: a user u is got by analyzing datasets and sampling joint probability distribution, and then distribution probability of u called p(u) is computed. Through computational formula of joint probability distribution, user u and time t is computed by p(u,t)= p(t|u) p(u); Based on p(u,t), p(l|u,t) is generated by obtaining corresponding Location L. Combining p(l|u,t) with p(u,t), p(u,t,l) is computed by p(u,t,l)= p(l|u,t) p(u,t); Thereby, combining p(u,t,l) with (1), p(R|u,t,l) is computed by p(R|u,t,l)=p(R|u,t,L)= p(R|u,t,L)= α p(R|L)+(1- α)p(u,t), where α is a tunable weight and then the context awareness model CA could be built by p(u,t,l,R) which is computed by p(u,t,l,R)= p(R|u,t,l) p(u,t,l). Generative process of the model could also be depicted by figure 1.

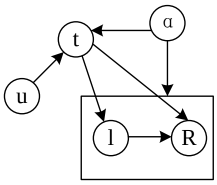

Figure 1. Context-awareness Model

1. Sampling the datasets and extracting the training sets u, $\{u_1,u_2,u_3\ldots,u_n\}$, which contains totally n users; Corresponding time sets $\{ t_1,t_2,t_3\ldots,t_n \}$ could also be extracted. Combining the two sets, the joint probability distribution $\{p(t_i|u_i),1\leq i\leq n\}$ is computed.

2. For every region r=1,…,R: ①a region r is extracted by standardized Gaussian dis-

tribution; ②the sets of GPS coordinates L={ {$l_{i,0}$, $l_{i,1}$}, 1≤i≤n } is obtained in a region r; ③Through (1), joint probability distribution of every region R is computed by $p(R|L)=\sum_{i=1}^{n}\sum_{r=1}^{R}p(r|(l_{i,0},l_{i,1}))$;

3. Matching GPS coordinates with nearby POIs. The rule is that comparing $p(R|L)$ with the number of POIs in region r. If $p(R|L)$ is greater than the number of points, the GPS coordinates in region are as a point matching with the nearest POI; On the contrary, the GPS coordinate which is closest to central point of a region would be matched with the nearest POI;

4. Through weight α, p(t|u) and p(R|L) could be tuned.

Referring to aforementioned procedures, the joint probability distribution of context-aware model is built.

2.4 Computing of Context-awareness Model

Through Gibbs sampling [13], parameters of context-aware model is estimated. Latent variable of the model is zumv, whose posterior probability distribution could be computed by (2).

$$P(z_{umv}|u_m,m_i,i_v,v_f)=\frac{P(z_{umv}|u,t,l,m,i,v)}{P(z_{umv}|u,t,l,m)} \quad (2)$$

Where u_m shows that tweet m posted by user u; m_i indicates that a clause i in tweet m; i_v manifests a word in clause I; v_f describes word frequency of a word v. $P(z_{umv}|u,t,l,m,i,v)$ represents joint probability distribution of z_{umv} expressed by some words v included in i clauses. These clauses comprise tweet m posted by user u at time t and in a position point l. $P(z_{umv}|u,t,l,m)$ describes that joint probability distribution of z_{umv} expressed by tweet m posted by user u at time t and in a position point l. So (3) and (4) could be computed by (2),

$$P(z_{umv}|u,t,l,m,i,v)=\frac{p(z_m,u_{t,l},m_{ivf})}{p(u,t,l,m,i,v)} \quad (3)$$

$$P(z_{umv}|u,t,l,m)=\frac{p(z_m,u_{tl})}{p(u,t,l,m)} \quad (4)$$

Where $p(z_m,u_{t,l},m_{ivf})$ presents distribution probability of topic z generated by tweet m which is posted by user u at time t in location l. m contains i clauses that include v words whose word frequency is f. $p(u,t,l,m,i,v)$ is joint probability distribution of i clauses and v words which include in tweet m posted by user at time t in location l. $p(z_m,u_{tl})$ is distribution probability of topic z generated by tweet m which is posted by user u at time t in location l, but clause number and word number of m is unknown. $p(u,t,l,m)$ is probability of tweet m posted by user u at time t in location l. z_m represents topic z that included in tweet m; $u_{t,l}$ represents user u at time t may be in location l; m_{ivf} represents word frequency f of word v which included in i clauses that those words comprise tweet m.

Based on aforementioned parameters, the purpose of perceiving context data could be attained. Next, posterior probability of (2) is ranked for POIs recommendation. In fact, greater posterior probability is, more ability of context-awareness is.

The relationship of user group and marked tweet number shows in Table 1. Due to recommendation results completed by users, so the results inevitably have highly subjective. Thus Cohen' Kappa coefficient is adopted for agreement to show reasonability of the results. The formula for calculating is $k=\frac{\Pr(a)-\Pr(e)}{1-\Pr(e)}$. In figure 2, most of K value is between 0.6 and 0.8, and k (2, 3) =0.821 is between 0.8 and 1. Because $k\in(0.6,0.8]$ achieves medium consistency and $k \in (0.8,1]$ achieves perfect consistency, so evaluation result of the four users is professional and precise.

Table 1. User Group and Marked Tweet Number

User Group Id	1	2	3	4
Marked Tweet number	20,000	20,000	20,000	20,000

Table 2. Correlation of Two Users and K

Correlation of two users	1,2	1,3	1,4	2,3	2,4	3,4
k	0.701	0.782	0.736	0.821	0.681	0.768

3. Experiments

3.1. Datasets

Experiments employ two data sets: Geo-Text and Sina-Tweets. Geo-Text is a public data set and Sina-Tweets is crawled by Sina Micro-blog API from Dec.1, 2012 to July 31, 2013. Details description is shown in Table 3.

Table 3. Datasets Description

Dataset name	Data Items	number	description
Geo-Text	User ID	9,481	User number
	Check-in time	377，616	Check-in time
	GPS coordinates	54，325	Location information
	Tweets	377,616	Micro-blog number
Sina-Tweets	User ID	201,066	User number
	Check-in time	197,627	Check-in time
	GPS coordinates	61,563	Location information
	Tweets	199,508	Micro-blog number

3.2. Evaluation Metrics

Recommendation Satisfaction Index (RSI), which is used to measure user's evaluation for recommendation. It is described the ratio of satisfactory recommendation (SP) and entire recommendation (ST) in (5),

$$RSI_j(i) = \frac{SP_{ij}}{ST_j}, \quad RSI = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n} RSI_j(i)}{\sum_{j=1}^{m} ST_j} \quad (5)$$

Where ,SP_{ij} is the number of people which expresses satisfaction with recommendation i in test set j; ST_j is the number of people in test set j; m and i is respectively the total number of test set and recommendation. $RSI_j(i)$ is recommendation satisfaction index of users express satisfaction with recommendation i in test set j and RSI is the mean value of $RSI_j(i)$.

3.3. Experimental Results and Discussion

In this subsection, Context Awareness model (short for CA) is compared with LCARS model and STT model, which are state of the art methods. In Figure 2, it is shown the incidence relation of tweet number, topic number, clause number and word frequency.

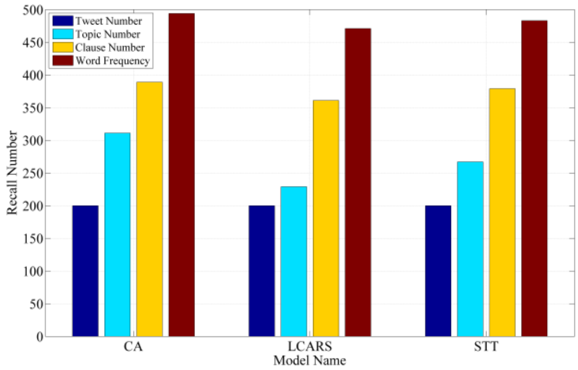

Figure 2. The Comparison of Recall Number on CA, LCARS and STT

In the above Figure, under the condition of same tweets, CA model has the best recall performance on topic number, clause number and word frequency. STT model comes second and LCARS comes last.

RSI describes the ratio of satisfactory recommendation (SP) and entire recommendation. In Figure 3, it is shown the experimental results of CA, LCARS and STT on the datasets of Geo-Text and Sina-Tweets.

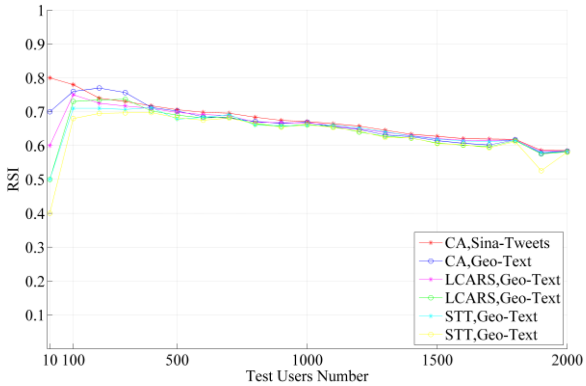

Figure 3. Comparison Results of RSI among CA, LCARS and STT

In the above Figure, when test users number of the three models is between 10 and 100, RSI rises rapidly to the top. When test users number continually increases, RSI drops gradually to 0.5. The results show that when test users number is small, RSI is strongly affected by the subjective factors. Under the condition of test user number rising, RSI gradually rises. However, when it rises to 1500, RSI is half satisfaction and half dissatisfaction. Comparison results of RSI among CA, LCARS and STT on datasets of Geo-Text and Sina-Tweets show that CA is much better fit to user’s actual requirements and LCARS or STT is next-best than CA model at the aspect of RSI.

Thus it could be seen that CA model has great improvement on recall performance and RSI through combining the advantages of LCARS and STT.

4. Conclusions

In this paper, a context awareness model called CA is present and in the model either spatial-temporal factors or the historic behavior of users are considered. In the experiments, Recommendation Satisfaction Index (RSI) of CA is evaluated. Experimental results show that CA has great improvement on accuracy rate of POIs recommendation and RSI compared with state of the art methods (LCARS and STT). In the future research, other idea on machine learning and artificial intelligence is addressed for improving accuracy of POIs recommendation to complete existing research, such as knowledge base *etc*.

Acknowledgements

The authors thank National Natural Science Foundation of China (61272109) and the Ph.D. Student Independent Scientific Research Project of Wuhan University (2011102110060).

References

[1]. H. Yin, B. Cui, Y. Sun, Z. Hu and L. Chen, "LCARS: A Location-Content-Aware Recommender System", KDD'13, **(2013)**.
[2]. B. Hu, M. Jamali and M. Ester, "Spatio-Temporal Topic Modeling in Mobile Social Media for Location Recommendation", ICDM, **(2013)**, pp. 1073-1078.
[3]. Q. Yuan, G. Cong, Z. Ma, A. Sun and N. Magnenat-Thalmann, "Who, Where, When and What: Discover Spatio-Temporal Topics for Twitter Users", KDD'13, **(2013)**, pp. 605-613.
[4]. J. Cohen, "A coefficient of agreement for nominal scales", Educational and Psychological Measurement, vol. 20, no. 1, **(1967)**, pp. 37-46.
[5]. J. Cohen, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit", Psychological Bulletin, vol. 70, no. 4, **(1968)**, pp. 213-220.
[6]. I. Konstas, V. Stathopoulos and J. M. Jose, "On social networks and collaborative recommendation", In SIGIR, **(2009)**.
[7]. K. W.-T. Leung, D. L. Lee and W.-C. Lee, "CLR: a collaborative location recommendation framework based on co-clustering", In SIGIR, **(2012)**.
[8]. J. J. Levandoski, M. Sarwat, A. Eldawy and M. F. Mokbel, "Lars: A location-aware recommender system", In ICDE, **(2012)**.
[9]. Y. Zhang, Q. Guo, L. Zhang and J. Lv, "Deep Web and Cognitive computing", Communications of CCF (China Computer Federation), vol. 10, no. 2, **(2014),** pp. 26-32.
[10]. B. Fu, J. Lin and L. Li, "Why People Hate Your App-Making Sense of User Feedback in a Mobile App Store", KDD'13, **(2013)**.
[11]. Y. Koren, "Factorization meets the neighborhood: a multi-faceted collaborative filtering model", In KDD, **(2008)**, pp. 426-434.
[12]. H. Yin, B. Cui, J. Li, J. Yao and C. Chen, "Challenging the long tail recommendation", Proc. VLDB Endow, vol. 5, no. 9, **(2012)**, pp. 896-907.
[13]. D. Venugopal and V. Gogate, "GiSS: Combining Gibbs Sampling and Sample Search for Inference in Mixed Probabilistic and Deterministic Graphical Models", AAAI, **(2013)**.

Authors

Fung Wong was born in 1984. He is a Ph.D. candidate at School of Computer, Wuhan University, and the student member of CCF. His research interests include social network and social media.

Shijun Lee was born in 1964. He is a professor at School of Computer, Wuhan University, and the senior member of CCF. His research interests include Web search and mining, the management and optimization of mass data, data mining, database technology and Android mobile development, *etc.*

Quanrui WONG, (1981) is a lecturer in the School of Information Engineer, Henan Institute of Science and Technology. She received BS degree in computer science and technology from Henan Normal University, Xinxiang, Henan province of China in 2004. And in 2007, she received her MS degree in computer software and theory from Henan Normal University, Xinxiang, Henan province of China. Now, she is a doctoral student in information engineer, Wuhan University of Technology, Wuhan, Hubei province of China. Her research interests are in data mining, analysis of users in social network and rough logic.

