

FCM: A Fine-Grained Crowdsourcing Model Based on Ontology in Crowd-Sensing

Jian An, Ruobiao Wu, Lele Xiang, Xiaolin Gui, Zhenlong Peng

► **To cite this version:**

Jian An, Ruobiao Wu, Lele Xiang, Xiaolin Gui, Zhenlong Peng. FCM: A Fine-Grained Crowdsourcing Model Based on Ontology in Crowd-Sensing. 13th IFIP International Conference on Network and Parallel Computing (NPC), Oct 2016, Xi'an, China. pp.172-179, 10.1007/978-3-319-47099-3_14. hal-01648003

HAL Id: hal-01648003

<https://hal.inria.fr/hal-01648003>

Submitted on 24 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



FCM : A Fine-grained Crowdsourcing Model based on Ontology in Crowd-Sensing

Jian An^{1,2}, Ruobiao Wu¹, Lele Xiang¹, Xiaolin Gui^{1,2}, Zhenlong Peng^{1,3}

¹ School of Electronics and Information Engineering, Xi'an Jiaotong University.
No.28, Xianning West Road, Xi'an, 710049, P.R. China

² Shaanxi Province Key Laboratory of Computer Network.
No.28, Xianning West Road, Xi'an, 710049, P.R. China

³ School of Business and Information Technology, Quanzhou Normal University.
Donghai, Quanzhou, 362000, P.R. China

{anjian@mail, wuruobiao@stu, xianglele@stu, xlgui@mail, msy@mail,
pengzl@stu}.xjtu.edu.cn

Abstract. Crowd sensing between users with smart mobile devices is a new trend of development in Internet. In order to recommend the suitable service providers for crowd sensing requests, this paper presents a Fine-grained Crowdsourcing Model (FCM) based on Ontology theory that helps users to select appropriate service providers. First, the characteristic properties which extracted from the service request will be compared with the service provider based on ontology triple. Second, recommendation index of each service provider is calculated through similarity analysis and cluster analysis. Finally, the service decision tree is proposed to predict and recommend appropriate candidate users to participate in crowd sensing service. Experimental results show that this method provides more accurate recommendation than present recommendation systems and consumes less time to find the service provider through clustering algorithm.

1 Introduction

In the new trend of ubiquitous computing, Crowd Sensing (CS) has been proposed to efficiently collect information in the multiple-source and heterogeneous environment. Based on massive information, including users' behavior characteristics, context information and activities law, the user behavior model and the relationship between service attributes could be obtained.

However, there are still some problems needed to be solved in the crowd sensing scenario. To start with, present crowdsourcing models do not particularly classifying the service requests [1]. Furthermore, the internal relationship between service preference of customers and historical data has not been fully considered in the existing service recommending model [2]. Thus, finding an applicable model to utilize the historical data is an effective way to promote success rate of service assignment.

To solve the problems mentioned above, a fine-grained crowdsourcing model has been proposed in this paper to realize sensing task assignment, which could recommend a set of appropriate service providers to the requesters. The whole system model is showed in the Fig.1. The key contributions of this paper are listed as following.

- In order to realize the reasonable distribution of sensing tasks, ontology theory is used in this paper to describe various attributes of users. Based on the ontology method, all service providers and requesters could be accurately differentiate.
- According to the fine granularity description of sensing requests, recommendation index of each service provider is calculated through similarity analysis.
- In addition, as a method to obtain approximate discrete valued target function, decision tree is also referred in the crowdsourcing model to predict potential attribute with the information entropy principle.

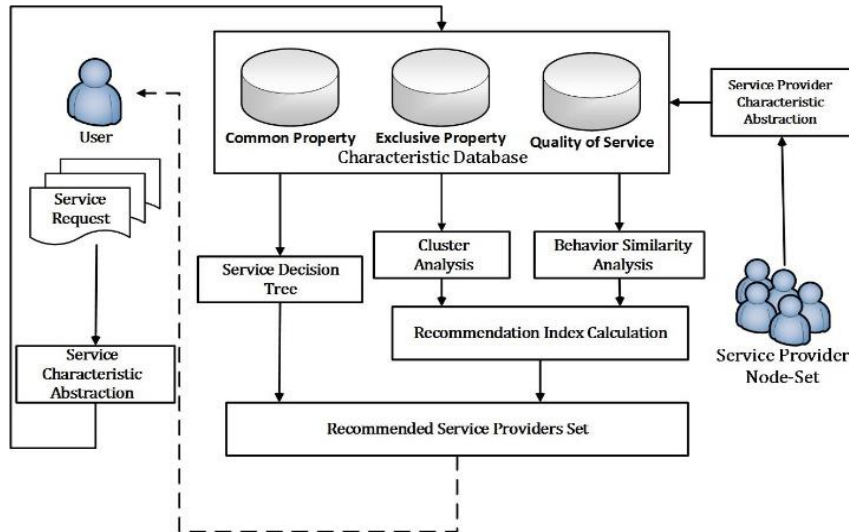


Fig. 1. Structure of service crowdsourcing model and whole process of recommending.

2 Characteristic Ontology Triple

The characteristic data should be defined in a unified form which could accurately describe the user behavior. Ontology, a concept data structure, could effectively show the user characteristic through concept sets and concept relationship set. In order to meticulously classify all service providers and customers, the characteristic is supposed to be defined as ontology triple $\langle CP, EP, QoS \rangle$. The components in each type of property are listed as following.

Common Property. For a service provider or a customer, common property contains user ID, user name and additional information. Therefore, the common property is defined as a set,

$$CP = \{ ID, Name, Additional Information \}$$

Particularly, all the elements of CP are text attributes, even though ID is the serial number of a service provider or a customer.

Exclusive Property. The exclusive property contains the location, availability and the service type. The Exclusive Property is defined as following set,

$$EP = \{ Location, Availability, Type \}$$

The element Location \in EP is the GPS location of the service provider or the customer. The element Availability \in EP describes whether the user could participate in the service. The type service providers specializing in or customers requesting for is selected from the type catalogue set which is determined subjectively in relation to the application scenarios.

QoS Property. Although the exclusive properties will affect the decision of users, the Quality of Service properties are the determinant in service selection. The QoS property is defined as following set,

$$QoS = \{ Valuation, Success Rate \}$$

Both of Valuation \in QoS and Success Rate \in QoS are $N \times M$ matrixs, where N is the number of service providers and M is the number of customers. Every element in the Valuation matrix is valued in the range of $[0,10]$.

3 Recommendation Method

3.1 Recommendation Index

The recommendation index shows selection priority of a service provider. The higher recommendation index of a service provider, the more likely the user will enjoy the service.

Definition 1 (User Behavior Similarity). In the service selection method, the user behavior similarity between two users is calculated by valuation offered in the past service. The user behavior similarity between two users U_i and U_j could be calculated through following equation,

$$Sim_{user}(U_i, U_j) = \frac{\sum_{SP_k \in OS(U_i, U_j)} (Val(i, k) - \overline{Val}_i)(Val(j, k) - \overline{Val}_j)}{\sqrt{\sum_{SP_k \in OS(U_i, U_j)} (Val(i, k) - \overline{Val}_i)^2 \cdot \sum_{SP_k \in OS(U_i, U_j)} (Val(j, k) - \overline{Val}_j)^2}} \quad (1)$$

Where $OS(U_i, U_j)$ consists of the service providers which have both served users U_i and U_j , $Val(i, k) \in [0, 10]$ is the valuation that user U_i values on the service provider SP_k .

Except for User Behavior Similarity, the types of service the service providers specialize in will also impact the validity of the valuation. Concept Hierarchy Tree [3] is proposed to processing the service types.

According to the definition of the information content in the information theory [4], the present information content of service type $Type_x$ will be calculated through the method proposed by the Resnik, and the calculate method is mentioned as following,

$$IC(Type_x) = -\log p(Type_x) \quad (2)$$

Definition 2 (Type Similarity). After obtaining the information content of each type, the type similarity $Sim_{sp}(Type_i, Type_j)$ between $Type_i$ and $Type_j$ will use the following formula to calculate,

$$Sim_{sp}(Type_i, Type_j) = \frac{IC(A(Type_i, Type_j))}{IC(Type_i) + IC(Type_j)} + \frac{d(A(Type_i, Type_j))}{2(d(Type_i) + d(Type_j) - d(A(Type_i, Type_j)))} \quad (3)$$

Where $A(Type_i, Type_j)$ is the nearest common ancestor concept of $Type_i$ and $Type_j$. $d(Type_i)$, $d(Type_j)$ and $d(A(Type_i, Type_j))$ are the depth of the $Type_i$, $Type_j$ and the nearest common ancestor concept of $Type_i$ and $Type_j$.

Recommendation index. Using user behavior similarity and type similarity, the recommendation index of service provider SP_j to user U_i is calculated as following,

$$RI(U_i, SP_j, Type_k) = Sim_{sp}(Type(SP_j), Type_k) \sum_{x=1}^n (Sim_{user}(U_i, U_x) \times Val(i, j)) \quad (4)$$

Where $Type(SP_j)$ is the Type property concluded in the exclusive property of SP_j , $Type_k$ is the service type which user requires, and n is the number of the users.

3.2 Service Decision Tree

Decision tree is an appropriate prediction model used in the process of selecting service provider. The attributes considered in the service decision tree conclude following data: Availability, Distance, Type and Success Rate.

After building the basis of the service decision tree by using the Ontology Triple, the rest of the service decision tree will be built through recursive process described as following step:

Step 1. Acquiring the data of node N_x which is going to build the its subtree, then defining the sample set remained as S_x , the attributes remained in the attribute list as $AttList_x$

Step 2. Check whether the attributes Valuation of samples in set S_x are the same. If the valuation attributes of samples in S_x are the same, this node should be marked as the leaves of the service decision tree. Conversely, if not all the valuation of samples is identical, turn to step 3.

Step 3. Check the attribute list. If the attribute list is empty, marking the major valuation attributes of the samples in S_x as the leaves of the service decision tree. Conversely, if the attribute list is not empty, turn to step 4.

Step 4. Find the most appropriate attribute $Attribute_x \in AttList_x$ as the decision factor and divide the sample into different child nodes of N_x and remove $Attribute_x$ from $AttList_x$.

Step 5. Recur Step 1 with the child node $N_x^*(S_x^*, AttList_x^*)$, S_x^* and $AttList_x^*$ are the sample set and attribute list in the node N_x^* .

4 Experiment

In the experiment, we have tested the fine-grained crowdsourcing model through data set collected from www.dianping.com. The data collects from 5000 customers and 100 service providers.

To show the recommendation process choosing the suitable service provider and improving the expect evaluation. In this experiment, Random Service Selection (RSS) and Convergent Population Diversity Handling Genetic Algorithm (CoDiGA) [5] are chosen to compare with Fine-grained Crowdsourcing Model (FCM). The experiment result shows in the Fig.2.

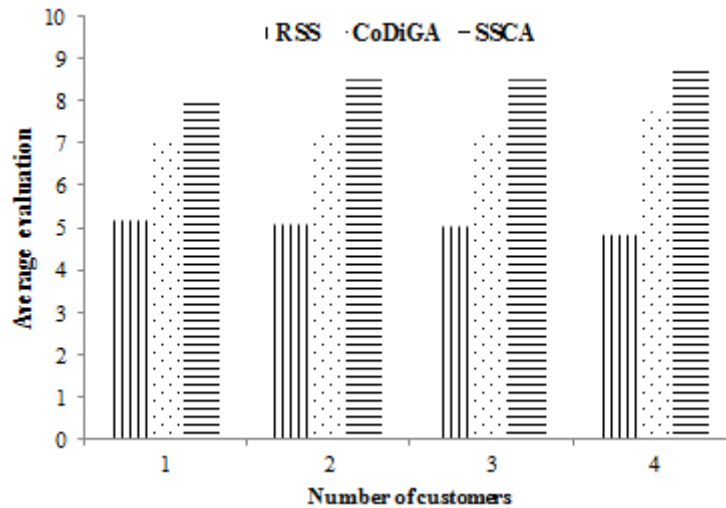


Fig. 2. Customers' average evaluation on recommended service providers after using RSS, CoDiGA and FCM in different sets of customers.

The recommendation effect of service decision tree should also be compared with recommendation index in the experiment. Hence, we choosing different numbers of service providers which have the highest prediction value gained from decision tree as the recommended service providers, in the meantime, choosing the same amount of service providers owns the highest recommendation index and choosing the same amount of service providers received the highest average valuation. The experiment result shows in the Fig.3.

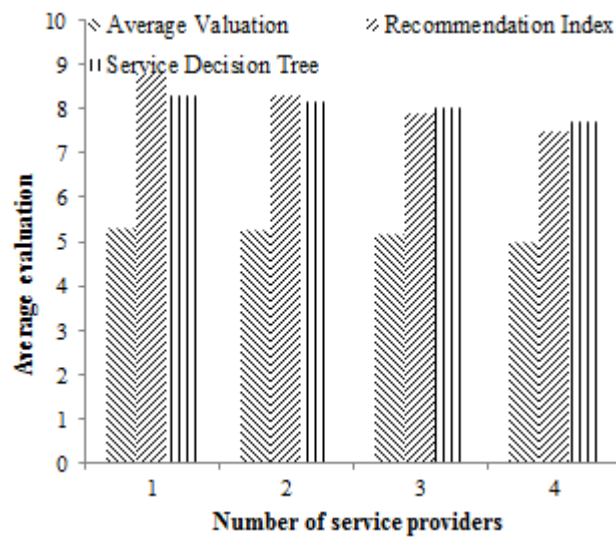


Fig. 3. The average scores valued by customers after accepting different number of recommended service providers through different recommendation methods.

In the fine-grained crowdsourcing model, different coefficient could lead to different results. The number of the customers, the number of the service providers and the number of clusters the server preparing to divide will impact the effectiveness of model. In order to affirm this model could be efficient in various environments. The experiment applies the fine-grained crowdsourcing model in different scales of customers and service providers, and gains the average valuation that customers score on the recommended service providers. The relationship between the scale of data and the average evaluation is showed in the Fig.4-Fig.5.

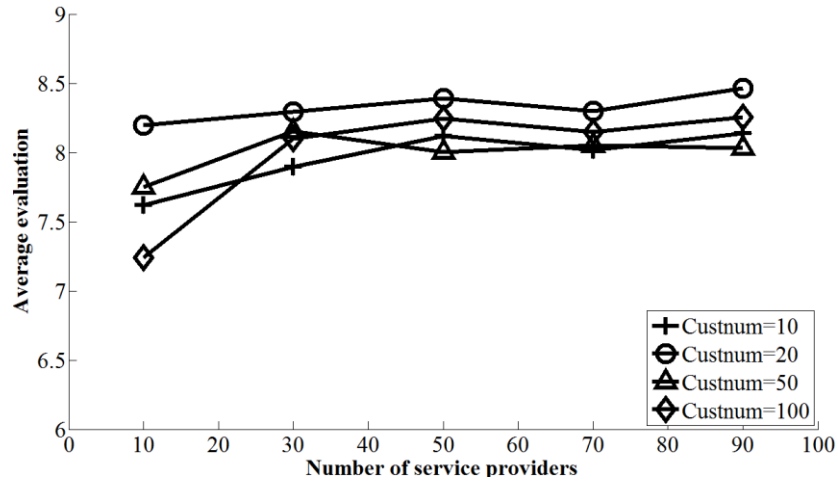


Fig.4. The effect of model on different data scales (M=2)

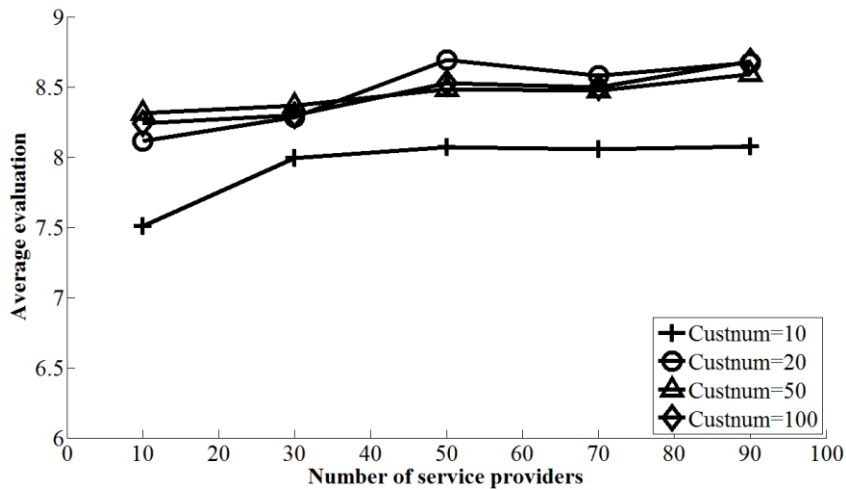


Fig.5. The effect of model on different data scales (M=5)

5 Conclusion

Taking the advantage of ontology theory and decision tree in the service recommendation could highly increase the customer satisfaction. The fine-grained crowdsourcing model discussed in this paper offering a good way to choose appropriate service providers.

In addition, we can find that if user needs a large amount of service provider, the service decision tree performs better, and if only a few service providers are needed, according to the recommendation index is a better choice. Both using service decision tree and recommendation index in the service recommendation could bring a perfect result to user.

Acknowledgments. This study was partly supported by NSFC under Grant (No.61502380, No. 61472316), Natural Science Basic Research Plan in Shaanxi Province of China (No.2014JQ8322), The Fundamental Research Funds for the Central Universities (No.XJJ2014049), and Shaanxi Science and Technology Innovation Project (2013SZS16-Z01/P01/K01), The Key projects of science and technology plan in Fujian province (No.2016H0029), Projects for Young Teacher in Fujian province (No.JA15401).

References

1. Doan, A., Ramakrishnan, R., Halevy, A. Y.: Crowdsourcing systems on the World-Wide Web. *Commun. ACM.* 54(4), 86-96 (2011)
2. Yang, W. S., Cheng, H. C., Dia, J. B.: A location-aware recommender system for mobile shopping environments. *Expert. Syst. Appl.* 34(1), 437-445 (2014)
3. Ye, F., Xiao, J.: Research on a data mining algorithm based on concept hierarchy tree and its application in CRM. In: *Proceedings of the International Conference on Computer Supported Cooperative Work in Design*, 694-697 (2004)
4. Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In: *Proceedings of the International Conference on Artificial Intelligence*, 448-453 (2010)
5. Ma, Y., Zhang, C. W.: Quick convergence of genetic algorithm for QoS-driven web service selection. *Compu. Netw.* 52(5), 1093-1104 (2008)