Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2016 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-4-2016

Ability Discovery and Weak Centralized Based Crowdsourcing Service Release System in Social Network

Zhenlong Peng
Xi'an JiaoTong University, Quanzhou Normal University, China, jxndpzl@163.com

Xiaolin Gui
Xi'an JiaoTong University, China, xlgui@mail.xjtu.edu.cn

Jian An
Xi'an JiaoTong University, China, anjian@mail.xjtu.edu.cn

Yali Ji
Xi'an JiaoTong University, China, jiyali@stu.xjtu.edu.cn

Ability Discovery and Weak Centralized Based Crowdsourcing Service Release System in Social Network

Peng Zhenlong, Xi'an JiaoTong University, Quanzhou Normal University, China , jxndpzl@163.com
Gui Xiaolin, Xi'an JiaoTong University, China, xlgui@mail.xjtu.edu.cn
An Jian, Xi'an JiaoTong University, China，anjian@mail.xjtu.edu.cn
Ji Yali, Xi'an JiaoTong University, China, jiyali@stu.xjtu.edu.cn

ABSTRACT

Crowdsourcing developed rapidly for its inspiring public abilities. But how to effectively find qualified participants and how to find and prevent malicious workers may be the main difficulties to ensure the crowdsourcing quality. In this paper, the related theories of social network were used in crowdsourcing services, the task publisher (Seeker) was regarded as the network center, his Abilities Set (AS) would be quantified and his Friends Abilities Matrix (FAM) would be generated according to the communication between them, thus his social network was re-constructed. Subsequently, some friends that conformed to the ability requirements of the task would be chosen to be the task receivers (Solvers). The natural trust relationship in the social network was fully used to build a crowdsourcing service release system on weak centralization. By using the social network, even the privacy information needn't to be shared with others, the system could help the seeker find solvers accurately in the seeker's own social network according to task demands, and then help to reduce fraud and invalid data. The simulation experiments showed that the release system could help the seeker discover his own abilities, construct the FAM, and select the appropriate solvers precisely and automatically.

Keywords: Social Network, Abilities Discovery, Crowdsourcing, Fog Computing, Friends Abilities Matrix

OVERVIEW OF CROWDSOURCING

Crowdsourcing refers to a company or organization outsources some work which should be done by employees in the past to non-specific (and usually large) network of the masses freely and voluntarily [1]. The idea of crowdsourcing got wide attention and was broadly used in more and more companies. Such as Uber, Amazon's Mechanical Turk, Drops A Car (in China) and so on. Crowdsourcing can integrate the mass wisdom to complete a task in a quick and low cost way, it embodies the collaboration spirit of **"**One For All, All For One**"**. It is seem as a new form of organizing productivity under the background of "Internet plus".

The main classifications of Crowdsourcing tasks

According to the data size of the uplink and downlink, the tasks can be divided into the following four types as shown in table 1. The downlink refers to releasing the tasks, while uplink refers to the collection of task results.

Table 1: Task classifications and typical applications according to data traffic

Type	Uplink traffic	Downlink traffic	Typical Application
I	*small*	*small*	*Questionnaire survey，Vote.*
II	*big*	*small*	*Collection of pictures and videos.*
III	*big*	*big*	*Compilation of pictures and videos.*
IV	*small*	*big*	*Label of pictures and videos*

Main frameworks of Crowdsourcing services

At present, the main architecture of crowdsourcing service is Centralization, namely the seekers and the solvers exchange data through a server, as shown in figure 1a. But the architecture will not work when the natural disasters such as earthquake, typhoon lead to malfunction in processing center, or when the information can't be sent to the processing center in the

wilderness. A no centralized routing strategy was proposed by Zhao Dong etc.[2], as shown in figure 1b. In the architecture, the tactic of **"Carrying—Storage--Foreword"** and the opportunistic meet between the mobile terminals were widely applied in this architecture to transfer tasks or bring the results back.

With the diversification of crowdsourcing tasks, the abilities of the participants become diversified and the number of the participant is also increasing. The type II, III or IV in table 1 will consume big data flow either for the seekers or the solvers. Especially, when the participants all use mobile network to release task by flooding way and the solvers are asked to transfer the results back instantly, a great data traffic cost will be generated, at the same time, the total energy consumption will increase greatly. No centralization framework tried to solve the class II and III in table 1, but the trust issues between the nodes is a big problem.

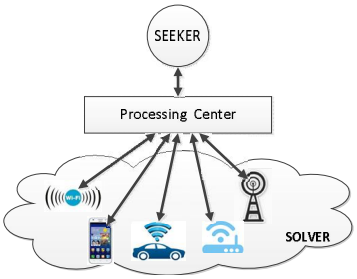


Figure 1a. Centralization

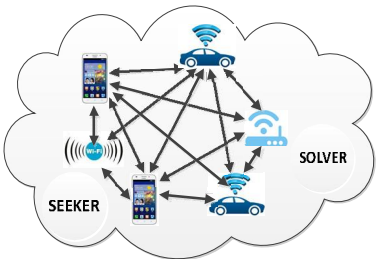


Figure1b. No Centralization

So we planned to use the social relations between nodes and build tasks distribution trust chain, but some global information sometimes need to be shared to each participating in the chain. Therefore, we proposed the weak centralization framework for crowdsourcing, as shown in figure 2.The weak centralization means few data will be sent to processing center except for "tasks/abilities classification, keywords table, honesty index, results evaluation", these information can be calculated and updated in a sampling time according to the demands. In fact, it is very difficult to obtain the social relations between friends and friends because of privacies; weak centralization framework is expected to avoid the problem in the largest extent. Just according to the interactive behaviors with his friends which he can get easily, each participant can re-construct the social network whose center is himself. The more suitable potential solvers are selected to execute or push forward the task, this framework is more feasible in reality.

In this framework, the fog computing [3-5] was introduced in. As the same as the no centralization, the seekers and the solvers were in the same set and they could communicate or collect data through the opportunity network. According to the technical characteristics of the fog computing, fog center could exchange the global data between the mobile sensors supervised by the fog in a certain region, as shown in figure 2, the fog 1 and fog 2 were able to handle or exchange the terminal data in their respective scope. Data exchange between fogs would be finished in the processing center. Then the load of processing center was greatly alleviated.

ABILITIES DISCOVERY AND TASKS RELEASE BASED ON SOCIAL NETWORK

In view of the above problems, we planned to use the social relations of the participants, their friend abilities can be found and quantitated automatically, and then the natural trust relationship in the social network was fully used.

Social Network

Social network is formed by some nodes and the connections between the nodes. The interaction between individual members within the social network will form relatively stable relations, and then influence people's social behaviors [6-9]. As shown in figure 4, A participant node can be denoted as, where means the attributes of the node. The social network can be denoted as, where means the friend nodes of central node ，
, where

Where means the strength degree between and , 0 means doesn't have the ability.

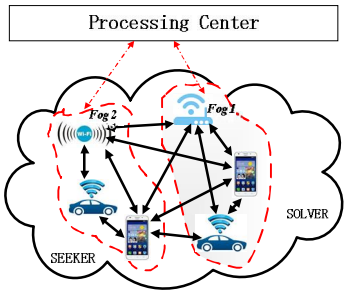


Figure 2:Week Centralization

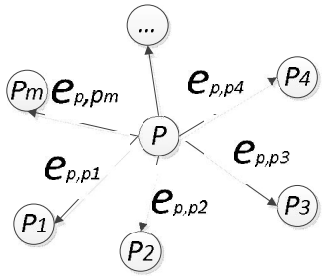


Figure3:Schematic Diagram of Social network

The abilities discovery based on content in social network

Each node has its own properties, such as hobbies, professional competence, etc., at the same time, the node usually has a judgment to the abilities of all his friends, such as what kind of friend were interested in what subject, what kind of friend was suitable to do what task, and so on. We agreed: was called as a node in social network and a task publisher in crowdsourcing system, had friends. Task publishers (Seekers) and task receivers (Solvers) were called participants.

Definition 1.[Abilities Set (AS)]. It is a set with all kinds of abilities to complete a crowdsourcing task. We agreed: the abilities set is in our system. AS was a global factor that should be shared in the processing center.

Definition 2.[Abilities Subset (ASS)]. It is composed by one or all of the elements from the AS.

Definition 3.[Abilities Value (AV)]. It is a digital denotation corresponding to the AS. We denoted it as. For example: means the abilities of node ， the original value is set by between 0 and 1, the default value is zero.

Definition 4.[Qualities Factor (QF)]. It is the comprehensive valuation given by all friends of a solver after the solver finished a crowdsourcing task. It can be denoted as. Hypothesis is the total task number that the friend invited to do，after completion, gives valuation according to the performance of every task, the valuation is marked , QF is denoted as :

where， means the weight of the task of the friend , .

Definition 5.[Honesty Index (HI)]. It is a weighted average of the QF evaluated by all of a solver's friends. We denoted it as , it is a global variable. For example, means the overall evaluation that all friends of node gave to ：

where， means the weight of the friend j to node ， generally, it can be set by 1.

Definition 6.[Friend Abilities Vector (FAV)]. A seeker, as the central node in his social network, gives the AVs for one of his friend based on the AS according to communications. For example, the FAV that node give to his friend can be denoted as:

Definition 7.[Friends Abilities Matrix (FAM)]. The *FAM* of a node is a matrix composed by all of the node's FAVs. For example, the FAM of node can be denoted as:

Re-constructing the Social Networks for the Participant

As discussed above, we can re-define the node in the social networks as a triple:
Where means the *HI*, means the *AVs*, means the *FAM*.

Computing and updating the AVs

（1）Normalization of the initial AVs for node :

（2）Dynamic updating AVs of

Algorithm 1. Dynamic Updating the AVs

```
Do while k<=n;
    Do while j<=m;
        …….(8)
  End Do;
End Do;
```

Where means the sampling times, means communication times between and his friend on ability . The AVs will update once time in every sampling period，and

Computing and updating the FAM

（1）Initialization of the FAM, it can be set by 0；

（2）Dynamic updating FAM of

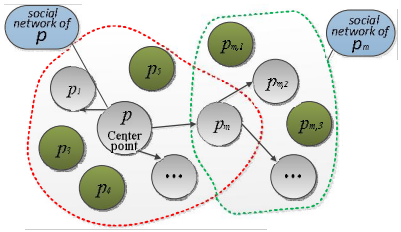
Algorithm 2. Dynamic Updating the FAM

Do while k<=*n;*
Do while j<=m;
…….(9)
End Do;
End Do;

where， means the total communication times between node and his friend ，and means the times on the ability/topic . As the formula 3 and algorithm 1 and 2 discussed above, node can re-construct itself as the following form:

Quick Task Release Mode Based Abilities Coverage.

After algorithm 3, the CNs of was determined, then could push the task forward to the CNs, as shown in Fig 4, the social network of was surrounded by the red dotted line, the CNs of maybe { so didn't push the task to { whose backgrounds were green. The friends received the task could complete the task, also could release the task in his social network by the same way. According to the "six degree separation"，theoretically, the task could be sent to anybody in the world by transferring 6 times [10-13]。Every time, we just let the participant push the task to friends in his social network,（the value of could be changed according to demands）. As a result, the release accuracy of the system would be higher than random release, so the friends could avoid interference by irrelevant information



The First Release The Second Release

Figure 4. The Task release mode based on social relationship

EMULATION

Wechat is the most popular mobile instant chat software, Until March 2016, the number of the monthly active users reached 762 million, 39% more than that of last year [14]. Our system is based on the Wechat data, at the same time, in order to avoid privacy leakage, the system only permits the users to analyze the contents and records of themselves.

Assumption: In the emulation, we roughly divided the tasks that were suitable for crowdsourcing in our lives into 10 types/abilities, there were denoted as A={1,2,…,10}. Node had 100 friends named from 1 to 100, the communication times between and his friends were between 5 and 500. At the first, several topics randomly were selected from the 10 topics; the probability density of the topics obeyed the Poisson Distribution. Then the FAM of was shown in table 2. Just 10 friends were included in the table because of the paper length, the number in the table means the communication times with different friends on different topics in a sampling cycle. In fact, this table is FAM of, it was also can be denoted as (formula 6 and 10).

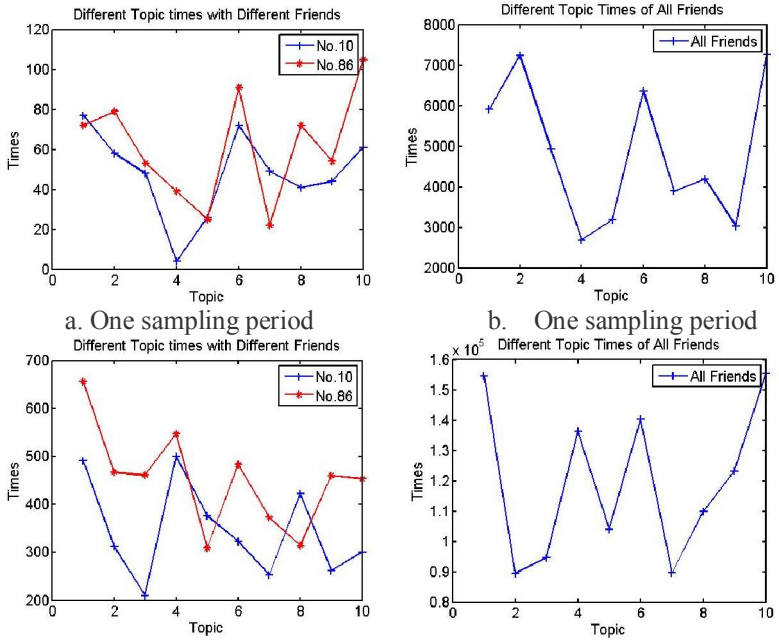
Table2:FAM of (10 friends)

Abilities / Friends	1	2	3	4	5	6	7	8	9	10
1	60	41	84	156	25	58	27	85	136	60
2	69	7	64	73	65	108	140	89	41	186
3	97	71	40	144	25	126	27	57	106	134
4	120	82	70	163	98	126	66	174	67	118
5	95	75	7	54	22	94	82	62	79	114
6	51	11	20	53	44	69	43	36	44	50
7	298	29	181	140	56	142	19	252	269	104
8	86	106	30	124	37	74	50	24	111	69
9	83	50	54	27	40	24	55	12	34	61

10	79	59	50	5	28	72	50	42	47	60

Abilities Discovery

The data in table 3 just covered one sampling period, in fact we also counted the communication times in 5 sampling periods. the AVs of could be calculated by algorithm 1, the differences between one sampling and 5 sampling were shown in fig 5, just No. 10 and No. 86 friends were randomly selected to do statistic.



a. One sampling period b. One sampling period

c. Five sampling periods. d. Five sampling periods

Figure 5:Communication times on different topics of friends

As the 4 figures shown above, we could find that fig a was similar to fig c, and fig b was similar to fig d. It implied that AVs of is relatively stable, and the node would like to communicate the topic that he was interested in.

CNs Selection

As a seeker to release a crowdsourcing task, should select ASS and the weight of ASS in the system. In the experiment, ability 7 and 10 were selected to generate a task respectively, and the CNs can be found by algorithm 3. The results were shown in table 3.

Table 3:The Selection result of CNs with single ability

Topic/ability No.	CNs
7	（42，65，82，12，60，71，36，88，30，59）
10	（12，34，51，88，62，74，95，17，70，65）

The simulation showed that the system could successfully count the communication times according to the AS, and calculate the AVs and FAM, at the same time they could be updated according to the sampling period. To any task generated in the system, the center of the network could find out the CNs correctly by matching the abilities demands and his friends. The CNs could repeat this procedure not more than 6 times, and then the task could be pushed forward rapidly, widely and precisely.

RELATED WORK

Crowdsourcing has attracted wide attention in the world since it was proposed about 10 years ago. Yuen etc. [15] concluded the research progress of crowdsourcing from applications, algorithms, performances and database; Kittur etc. [16] expounded the challenges of the crowdsourcing from 12 aspects such as the principle, cooperative work, inner regulations, real-time response, primary applications etc., and then indicated the development direction in the future. Shamir etc. [17] analyzed the consistency of artificial classification and annotations, and a mothed was proposed to control the crowdsourcing quality by supervised machine learning; Zhao etc. [18] summarized crowdsourcing from information, technologies, public and organization, and then indicated the deficiencies and the possible development objects. On the aspect of task transfer，considering the mobility of the mobile terminals, Tuncay etc. [19] proposed opportunistic encounter to release task in a distributed way; Hua (Jonathan) Ye [20] analyzed the crowdsourcing intentions based on cost theory and the resource-based view and validates using survey data from 161 organizations, the results indicated that the low cost, brand visibility, and access to specialized skills were the positive factors, but codification costs and evaluation costs were negative factors that influenced firms' crowdsourcing intentions.

CONCLUSION

We believe, participant coverage should be extended to the “location, attributes, background knowledge, social relations, and credibility”. The system can discover participants' abilities dynamically and automatically from the communication between the seeker and his friends by analyzing the content, time, records in his social network, and then the AVs , FAM , HI can be calculated out. So the social network of the seeker can be re-constructed. Using the natural trust relationship implied in the social relations, the system can push the task to the potential participants more precisely and quickly than release randomly, because people always prefer to complete and transfer the task from his social network. The simulation results verify the effectiveness of the system.

ACKNOWLEDGMENT

This research was supported by the funds from Fund project:

(1) Natural Science Foundation of China(No. 61172090,61472316,61502380)

(2) Key projects of science and technology plan in Fujian province , "Key techniques of task assignment and results

(3) collection on mobile crowd sensing based on trustable interaction "(No.2016H0029)

(4) Projects for Young Teacher in FuJian province 2015: “Research on Integration development of Mobile E-commerce and Inner location based on iBeacon”(No.JA15401).

(5) Key Projects of Quanzhou Technology Bureau in 2014: “Study on integration application and development mode for mobile E-commerce and the key technologies of IOT” (No.2014Z131).

(6) Natural Science Foundation of FuJian Province（No. 2015J01286）.

REFERENCES

[1] Howe J（2006）. The rise of crowdsourcing[J]. Wired Magazine,14(6):1-4;

[2] Zhao Dong，Ma, Huadong; Tang Shaojie,et al（2015）. A cooperative framework for building sensing maps in mobile opportunistic networks[J]. *IEEE Transactions on Parallel and Distributed Systems*, 26(2):392-402;

[3] Hongxing Li, Guochu Shou, Yihong Hu, et al（2016）. Mobile Edge Computing: Progress and Challenges[C]. *4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, 83-84;

[4] Eleonora Cau, Marius Corici, Paolo Bellavista, et al（2016）. Efficient Exploitation of Mobile Edge Computing for Virtualized 5G in EPC Architectures[C]. *2016 4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*,100-109;

[5] Shanhe Yi, Zijiang Hao, Zhengrui Qin, et al（2015）. Fog Computing: Platform and Applications[C].*2015 Third IEEE Workshop on Hot Topics in Web Systems and Technologies*,73-78;

[6] Martínez-Núñez, Margarita1, Pérez-Aguiar, Waldo Saul（2014）. Efficiency analysis of information technology and online social networks management: An integrated DEA-model assessment[J]. *Information and Management*, 51(6):712-725; DOI: 10.1016/j.im.2014.05.009;

[8] Lee Rainie, Barry Wellman．Networked（2012）: The New Social Operating System．London：*The MIT Press*，;

[7] Z˘iberna Aleš , Vehovar, Vasja（2009）,Using social network to predict the behavior of active members of online communities[C].*Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining,* 119-124.

[9] Alderson D., Doyle J.C. and Li L., et al (2005), Towards a theory of scale-free graphs: definition, properties, and implications[J]. *Internet Mathematics* 2(4):431-523,.

[10] Travers J, Milgram S(1969). *An experimental study of the small world problem* [J]. Sociometry. 32 (4): 425–443.

[11] WattsD,StrogatzS(1998). *Collective dynamics of‘small-world’networks*[J]. Nature. 393 (6684): 440–442.

[12] Newman M, Strogatz S, Watts D(2001). *Random graphs with arbitrary degree distributions and their applications* [J]. Physical Review E. 64 (2): 26118.

[13] Kleinberg J(2000). Navigation in a small world [J]. Nature., 406 (6798): 845.

[14] The monthly active user of WeChat surpassed 760 million, nearly catch up with QQ[EB/OL]. http://business.sohu.com/20160520/n450508941.shtml

[15] Yuen M,King I,Leung K(2011). A survey of crowdsourcing systems// *Proceedings of the IEEE 3rd International Conference on Social Computing.Boston,USA*,:766-773.

[16] Kittur A, Nickerson J V,Bernstein M S, et al(2013). The future of crowd work, Proceedings of the *2013 ACM Conference on Computer Supported Cooperative Work. San Antonio,USA*, 1301 -1318.

[17] Shamir, Lior, Diamond, Derek, et al(2015), Leveraging Pattern Recognition Consistency Estimation for Crowdsourcing Data Analysis, *IEEE Transactions on Human-Machine Systems,* (99):1-7

[18] Zhao Yu-Xiang,Zhu Qing-Hua(2012). Evaluation on Crowdsourcing reasearch:Current status and future direction. *Information Systems Frontiers*,11(1):1-18.

[19] G S Tuncay, G Benincasa, A Helmy(2012). Autonomous and distributed recruitment and data collection framework for opportunistic sensing [C]. *Proceedings of the 18th annual international conference on Mobile computing and networking, ACM MobiCom'12*, 407-409.

[20] Hua (Jonathan) Ye (2015), Atreyi Kankanhalli, Investigating the antecedents of organizational task crowdsourcing[J]. *Information & Management,* 52, 98–110;