

TASK ASSIGNMENT IN CROWD SOURCING USING VECTOR SPACE MODEL

A.Vijayalakshmi¹ & Prof.Chittranjan Hota²

¹Lecturer ,Department of Computer science, BITS- Pilani, Hyderabad.

²Professor, Computer science Department , BITS- Pilani, Hyderabad.

Abstract: Recently the crowd sourcing has become a well known model for doing different assignments in different field. But there are some challenges and issues. The important issues in crowd sourcing are the task assignment and quality. Assigning task to particular resource is a challenge as most of people are unfamiliar, unknown about crowd sourcing and maximum assignment is done online. We address this issue by finding the experts and assign the task based on their expertization. Here we assumed that expert people will always produce quality work. We used this technique in our social media application which is especially created for faculties .We have used a vector space model for finding experts which is normally used for retrieval of document based on the query in a search engines.

Key terms: Crowd sourcing, social media application, vector space model, expert finding, task assignment

1. Introduction

Crowd sourcing is the process of involving more people to do one task like label photos ,language translations, logo design ,getting ideas .The task is done by the people who are especially from online rather than from employees of any organization. The crowd sourcing techniques has been existed for many years for example face book has been using crowd sourcing technique for translation of their website into different language since 2008 and another example like Wikipedia which is written by people like professional to normal person in all over the world. Hence the crowd sourcing is in place since long, however after Jeff How mentioned in his article rise of crowd sourcing [1],it just becomes famous in various fields. Jeff defined it as 'the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined network of people in the form of an open call'[2]. The expression "crowd sourcing" is much the same as "outsourcing" in any case, the distinction is that in outsourcing the task is given to the authoritative specialist yet in crowd

sourcing the undertaking is allotted to individuals who are in online.

The crowd sourcing works in the following way. First the tasks should be identified and then the tasks should be assigned to online communities randomly those who are interested to do the task. The crowd may be any person who is willing to do the task. There are two types of crowd sourcing ie explicit crowd sourcing and implicit crowd sourcing.

Explicit crowd sourcing means the task has been assigned to workers directly but in implicit crowd sourcing, the answer for the specific issue will be found while doing some different undertakings by the user and without the knowledge of users.

Crowd sourcing relies on human laborers to finish work, yet people are tend to make errors and which can make the results bad. There are two reasons. First, some wicked worker can submit the random result for the assigned tasks, in order to get rewards. The output quality would be degraded by this attitude of workers and secondly, the worker may not have sufficient knowledge or skill set for completing the particular task. To address this issues, different methods are followed in many crowd sourcing platforms like in Amazon mechanical Turk ,single task is divided into multiple task and then it is assigned to multiple worker .After collecting the result ,it is again given to some other people to check the quality and some platform using majority voting method but using this method time and money would be wasted so we have tried a new method ie using vector space model method. The vector space model is used because of simplicity and efficiency over large document collection .The vector space model produces the list of workers based on their expertization.

2. Related work

For task assignment ,Chien ho jo and Jennifer Wortman Vaughan used Dual task assigner algorithm in which task will be assigned to the worker randomly[1].In this dual task assigner algorithm, it is assumed that initially workers skills are unknown, but once the workers complete the task their skill set

would be learned. Michal Kosinski proposed IQ test to solve the task assignment problem. Here questions were asked from widely used IQ test—Raven’s Standard Progressive Matrices (SPM)[2]. Alessandro Bozzon created a framework to assign the task [3] Ece Kamar used probabilistic model to predict the user behavior[4] and in bipartite graph is used to assign the task by David R Sewoong [5], spammer hammer algorithm and two phase exploration and exploitation algorithms are used to assign the task to the user in [6] [7]. In Quara, Voting method is used to find out the best quality content [8]. Our work differs from previous work in such a way that we will assign task based on the experization of that user. We used vector space model for assigning task. It is a one of the information retrieval models and normally used for selecting the document based on the query in the search engines. Vector space model requires much less labor than other approaches which used semantics such as ontology because they extract knowledge automatically from a given document [9].

3. Methodology of task assignment

We have created a social media application for faculties of multi campus universities to share their academic and research experience. We have used social media as a crowd sourcing platform because social media network is a only way to attract and connect the people around the world. Also a new report from the Babson survey research group and Pearson leading learning company finds that college faculty members have become sophisticated consumers of social media for their various needs. This social media consists of Registration details, Forums, Chat application, blogs, albums and also we have provided the platform for posting task and applying task. So the faculties can share their academic and research experiences and write their own blog and clarify other faculties doubt by posting queries on forum. Also they can post their photos as well.

Expert selection is an important aspect of many web applications especially on crowd sourcing because for some tasks like writing books or writing lecture material collaboratively, the knowledge of that particular task is important otherwise faculties cannot give the quality output. This was crucial for our current research on crowd sourcing, and therefore we dedicated a specific research line to this aspect

We have addressed this issue by assigning the task by ranking the faculties according to the knowledge that they have about the given task, after that the task will be assigned to the top experts. Our method is similar to Explicit semantic analysis (ESA) method used in [11] for expert finding. In ESA, Wikipedia is used to find

the semantic relatedness but we use vector space model because of this model has the advantage of partial matching. For that, first, we collect the data from various application like chat application forum and album and using vector space model we will find out the experts in particular topic. After finding experts we will assign the work to them, so that we will get reliable worker and good quality work. This new approach will improve the result of task assignment on crowd sourcing.

4. Pseudo code for task assignment

1. We have created faculties profile by collecting frequent used words from registration details, blogs, question forum and chat applications.
2. Calculated Inverse document frequency using the following formula

$$\text{IDF}(t) = \log(N/n_t) \tag{1}$$

Where N =total number of faculties profile we have, n_t =the number of faculties profile containing the word t.

3. Calculated term frequency

$$\text{TF}(t) = (n_t / N) \tag{2}$$

Where n_t = Number of times term t appears in a faculty profile, N=Total number of terms in the document.

4. Find idf*tf for each faculty
5. Find tf*idf values for the given task
6. Calculate the length of each faculties profile and given task
7. Find the cosine similarity for each faculty

$$\text{Cos}(\Theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i * B_i}{\sqrt{\sum_{i=1}^n A_i^2} * \sqrt{\sum_{i=1}^n B_i^2}} \tag{3}$$

Where A_i and B_i are components of vectors A and B ie faculty profile vector and task vector

8. Sort out the list of faculties in descending order

A: Example:

Expert finding using Vector space model

Consider a following 4 faculties profile represented as a no of keywords ie vectors

f1: (Computer networks(CN) Multimedia(MM) database(DB))

f2: (Computer networks(CN) Multimedia(MM) Operating system(OS))

f3: (Data ware housing(DW) data mining(DM) database(DB))

f4: (Data ware housing(DW) database(DB) Operating system(OS))

Some terms appear in three profiles, some appear only in one profile.

The total number of profiles are N=4.

The idf values for each terms in the profile:

Data warehousing $\log_2(4/2) = 1$

Data mining $\log_2(4/1) = 2$

Computer networks $\log_2(4/2) = 1$

Operating systems $\log_2(4/2) = 1$

Multimedia $\log_2(4/2) = 1$

database $\log_2(4/3) = 0.378$

Then we need to find the Tf scores for all the terms Tf scores are calculated and put it in the matrix form.

In the matrix, the keywords Data ware housing, Data mining, Computer networks, Operating system ,Multimedia and Data base are represented as DW,DM,CN,OS,MM and DB.

Table 1. Tf scores for all 4 faculties profiles

	DW	DM	CN	OS	MM	DB
F1	0	0	1	0	1	1
F2	0	0	1	1	1	0
F3	1	1	0	0	0	1
F4	1	0	0	1	0	1

Now we need to multiply tf and idf values of all faculties profiles

Table 2. Multiplication of idf and tf values

	DW	DM	CN	OS	MM	DB
F1	0	0	1	0	1	0.378
F2	0	0	1	1	1	0
F3	1	2	0	0	0	0.378
F4	1	0	0	1	0	0.378

Then we should represent the task as vectors .

Given the following task: (computer networks computer networks database),

Then we should calculate the tf-idf vector for the task as well. When computing the tf-idf values for the task terms, we divide the frequency by the maximum frequency (2) and multiply with the idf values.

Table 3. tf and idf values for the given task

	DW	D M	CN	OS	MM	DB
T	0	0	$(2/2)*1=1$	0	0	$(1/3)*0.378=0.126$

Length of each faculty Profile and task:

Length of f1 = $\sqrt{1^2+1^2+0.378^2} = 1.46$

Length of f2 = $\sqrt{1^2+1^2+1^2} = 1.732$

Length of f3 = $\sqrt{1^2+2^2+0.378^2} = 2.267$

Length of f4= $\sqrt{1^2+1^2+0.378^2} = 1.46$

Length of T = $\sqrt{1^2+0.126^2} = 1.003$

Then the similarity values are:

$\text{CosSim}(f1,T) = (0*0+0*0+1*1+0*0+0.126*0.378) / (1.46*1.003) = 0.717$

$\text{CosSim}(f2,T) = (0*0+0*0+1*1+1*0+1*0+0.126*0) / (1.732*1.003) = 0.575$

$\text{CosSim}(f3,T)=(1*0+2*0+1*0+1*0+0*0+0.378*0.126/ (2.267*1.003) = 0.0015$

$\text{CosSim}(f4,T)=(1*0+0*0+1*0+1*0+0*0+0.378*0.126/ (1.46*1.003) = 0.0325$

According to the similarity values, the final order in which the faculties are ranked as per the task is

: f1, f2, f4,f3

5. Experiment

A. Experimental setup:

We have created social media for faculties. The faculties are actively participated in all forum ,article creation and writing blogs . We have collected data from 15 faculties and created a profile for each faculty. We have posted some task on our social media application .All faculties applied for the given task using apply task facility We have already created a profile for each faculty which consists of some keywords ie vectors and then using tf and idf values of faculties and the given task we have found the similarity between the faculties and task, After finding the similarities we have ranked the faculties as per the similarity value. Finally we got the list of faculties ie list of experts as per the rank .

Also we have calculated the precision and recall value for the vector space model .Precision and recall are used as the basic measures for evaluating the relevance .So we have calculated this precision and recall value for vector space model and compared that model with language model.

Precision = A/A+C	-4
--------------------------	-----------

Where A =No of Relevant records received
 C =No of irrelevant records received

Recall =A/A+B	-5
----------------------	-----------

Where A=No of relevant records received
 B =No of relevant records not received

B. Experimental result:

The below table shows the performance of vector space model for the three tasks and we observed that the vector space model performs better than the language model, which is the another type of information retrieval algorithm, for task assignment in crowd sourcing.

Table 4. Performance of vector space model

Task	LM		VSM	
	Precision	Recall	Precision	Recall
Cloud Computing	0.4	0.1	0.6	0.27
Data mining	0.5	0.1	0.5	0.18
Operating system	0.6	0.27	0.8	0.32

6. Conclusion and Future work

Task assignment is the difficult task on crowd sourcing especially for heterogeneous tasks. A lot of algorithms and methods have been used on crowd sourcing platform to assign the task . In this paper we have proposed the new model ie vector space model for task assignment on crowd sourcing platform. Using this vector space model we got the list of experts with the rank then we assigned the task based on the rank. This is the first stage of our research. Further, we will be using trust algorithm to find out the degree of belief of each faculty. Because we cannot say that expert people are always trustable to solve any issues. So the trust algorithm will be applied on the list of experts obtained from vector space model to get the faculties who are knowledgeable and trustable to assign the task .

References

[1] Jeff Howe(Jan 06,2006) ,Rise of crowd sourcing, [online] Available: <http://www.wired.com/2006/06/crowds>

[2] J. Howe, "Crowd sourcing : Crowd sourcing: A Definition," Jun-2006. [Online]. Available: http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html. [Accessed: 02-Sep-2016].

[3] Ho, Chien-Ju and Vaughan, Jennifer Wortman. "Online task assignment in crowd sourcing markets".in AAAI, 2012 .

[4] M. Kosinski, Y. Bachrach, G. Kasneci, J. Van-Gael, and T. Graepel, "Crowd IQ: measuring the intelligence of crowd sourcing platforms," 2012, 151–160.

[5] Bozzon A, Catallo I, Ciceri E, Fraternali P, Martinenghi D, Tagliasacchi M, Tagliasacchi M "A framework for crowd sourced multimedia processing and querying", In: Crowd Search, (2012) pp 42–47.

[6] Ece Kamar, Severin Hacker, and Eric Horvitz. "Combining human and machine intelligence in large-scale crowd sourcing". In AAMAS, 2012.

- [7] D.R. Karger, S. Oh, D. Shah. "Budget-optimal crowd sourcing using low-rank matrix optimizations". In Proc. of Allerton Conference on Communication Control and Computing, 2011 .
- [8] Karger, D. R., Oh, S., and Shah, D. Iterative learning for reliable crowd sourcing systems. In Proceedings of NIPS, 2011.
- [9] Wang, G.; Gill, K.; Mohanlal, M.; Zheng, H. & Zhao, B. Y. " Wisdom in the social crowd: an analysis of quora", in 'WWW' , International World Wide Web Conferences , (2013), pp. 1341-1352.
- [10] Panagiotis G. Ipeirotis, Foster Provost, and Jing Wang. "Quality Management on Amazon Mechanical Turk," in Proceedings of the ACM SIGKDD Workshop on Human Computation (HCOMP'10), pp. 64-67, 2010.
- [11] P.D. Turney and P. Pantel." From frequency to meaning: vector space models of semantics", 2010, Journal of Artificial Intelligence Research, 37:141–188
- [12] Mark A.Greenwood ,"Implementing a vector space document retrieval system",Oct 2002.
- [13] Li, Xiu, et al. "A Service Mode of Expert Finding in Social Network." Service Sciences (ICSS), International Conference on. IEEE, 2013.
- [14] T. Padmapriya and V.Saminadan, "Handoff Decision for Multi-user Multiclass Traffic in MIMO-LTE-A Networks", 2nd International Conference on Intelligent Computing, Communication & Convergence (ICCC-2016) – Elsevier - PROCEEDIA OF COMPUTER SCIENCE, vol. 92, pp: 410-417, August 2016.
- [15] S.V.Manikanthan and D.Sugandhi " Interference Alignment Techniques For Mimo Multicell Based On Relay Interference Broadcast Channel " International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume- 7 ,Issue 1 –MARCH 2014.
- [16] Rajesh, M., and J. M. Gnanasekar. "Path observation-based physical routing protocol for wireless ad hoc networks." International Journal of Wireless and Mobile Computing 11.3 (2016): 244-257.

