



**HAL**  
open science

## An answerer recommender system exploiting collaboration in CQA services

Ning Wang, Marie-Helene Abel, Jean-Paul Barthès, Elsa Negre

► **To cite this version:**

Ning Wang, Marie-Helene Abel, Jean-Paul Barthès, Elsa Negre. An answerer recommender system exploiting collaboration in CQA services. IEEE International Conference on Computer Supported Cooperative Work in Design (CSCWD 2016), May 2016, Nanchang, China. pp.198-203, 10.1109/CSCWD.2016.7565988 . hal-01386789

**HAL Id: hal-01386789**

**<https://hal.science/hal-01386789>**

Submitted on 31 Oct 2016

**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.

# An Answerer Recommender System Exploiting Collaboration in CQA Services

Ning Wang, Marie-Hélène Abel, Jean-Paul Barthès  
Sorbonne Universités  
Université de Technologie de Compiègne  
CNRS UMR 7253 Heudiasyc  
Compiègne, France  
Email: {ning.wang, marie-helene.abel, barthes}@utc.fr

Elsa Negre  
Université Paris-Dauphine, CNRS,  
UMR 7243 Lamsade  
Paris, France  
Email: elsa.negre@dauphine.fr

**Abstract**—Community-based Question Answering (CQA) services are becoming popular as the public gets used to look for help and obtain information. Existing CQA services try to recommend someone for answering new questions. On the other hand, people are allowed to exchange information and experience using various collaborative tools. It would be interesting to combine the two approaches to increase the reliability of recommending an answerer. Thus, relying on semantically modeled traces, we propose a comprehensive approach that recommends an answerer in a collaborative environment. From a global point of view, this approach consists in evaluating users by the performance in the CQA services and the corresponding knowledge sharing activities in which they participated in a collaborative context. By modeling and analyzing users’ behavior, we assess the competency of an answerer in a particular collaborative context.

## I. INTRODUCTION

Community based Question Answering (CQA) services are defined as dedicated platforms allowing users to answer other users’ questions, resulting in the building of a community in which users share and interactively give ratings to questions and answers (Liu et al. [1]). Good examples such as Yahoo! Answers, Baidu Knows as well as more social-oriented newcomers as Quora<sup>1</sup> have gained popularity over the recent years [2]. In some cases when people need personal experience or real examples of applying knowledge, they prefer to seek an answer “with a human touch.” Answerers are usually attracted by less tangible incentives and social rewards such as reputation or points (Raban [3]). Millions of participants in CQA are mainly volunteers attracted by the openness and accessibility. A social network based on real identities encourages users to build reputation while access to user histories enables them to judge the reputation of others as discussed by Paul et al. [4]. At the same time, a social voting system which combines ranking algorithms for answers, allows participants to identify and promote high quality answers. For example, Zhihu, a famous Chinese CQA website which started in 2011, already has more than 10 million active users who have proposed and answered over 2 million various questions. Due to the large number of participants, it is difficult to find a right answerer to a newly-posed question, which leads to a large amount of questions unanswered. Thus, recommending an answerer for a new question needs to be achieved.

<sup>1</sup>Yahoo! Answers: answers.yahoo.com; Baidu Knows: zhidao.baidu.com; Quora: quora.com

It has been years that people work on the Internet in a collaborative context. They share documents by Dropbox and Google Drive. This gap between a collaborative context and CQA services can be bridged to increase the reliability of recommending an answerer. Thus, we propose an approach for recommending an answerer to a new question in a collaborative context that allows users to organize their resources and knowledge into a semantic web. We integrate the evaluation of traces of both CQA services and other services including sharing various types of documents, editing Wikipedia articles, etc., in order to recommend a competent answerer as soon as new questions arrive.

The remainder of this paper is organized as follows. In Section II we identify various merits and limitations of current studies on CQA services. Section III introduces the problem statement, previous work on semantic traces and our approach. In Section IV we propose a prototype of a semantic platform “MEMORAE” which integrates CQA services and an answerer recommender system by exploiting all traces. Section V gives conclusions and mentions directions for future work.

## II. BACKGROUND

Our work stretches across three areas of research: semantic collaborative applications, CQA services and recommender systems. In this section we first give an overview of nowadays most popular CQA sites. Then related recommending research work will be discussed.

### A. Interest of CQA Services

As the most successful pioneer of CQA service provider, Yahoo! Answers hosts over a billion answers on a wide variety of topics as mentioned by Liu et al. [2]. Typically the lifecycle of a question is as follows: an asker posts a new question and the system responds with a list of topics to index it, as shown in Figure 1. Meanwhile, each answer belonging to the index can be annotated and qualified by a vote of all participants, including the asker himself. These findings have important implications for the design of successful community knowledge-building sites. Finally, the system lines up all the answers according to how users have voted.

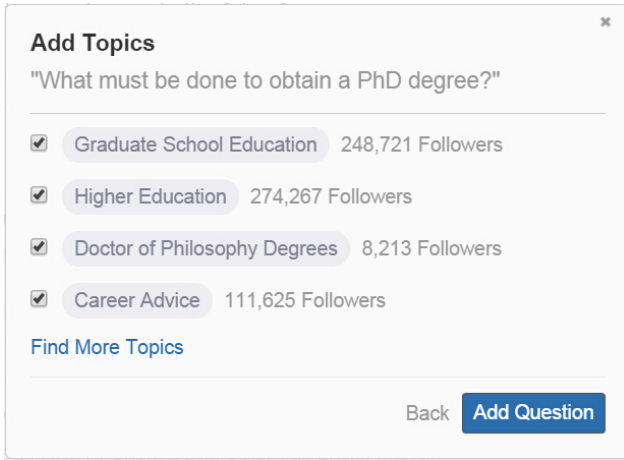


Fig. 1. The step of checking topics when posting a question in Quora.

### B. Related Work

Since the emergence of CQA services, ranking and recommending an answerer has been an active area of research. The recommendation process aims to guide users in their exploration of large quantities of data available by identifying relevant information as shown by Negre [5]. To provide relevant answers, significant research has been done. Liu et al. [6] combined language models with an analysis of user’s profiles and achieved satisfying performance. Agichtein et al. [7] worked on evaluating the quality of content in CQA. Shah and Pomerantz [8] predicted response success and quality. There is a corresponding problem of complex QA evaluation where recent efforts at automatic evaluation show that even for well-defined objective complex questions, evaluation is labor-intensive and poses many challenges as shown by Lin and Demner-Fushman [9], or Lin and Zhang [10].

Furthermore, it is attractive to route new questions to answerers to improve the answer quality. Jurczyk and Agichtein [11] proposed a graph structure for CQA systems to discover authoritative users in topical categories. Bouguessa et al. [12] focused on automatically discriminating between authoritative and non-authoritative users by modelling users authority scores for each topic. Additional question routing methods are discussed by Liu and Agichtein [13], Liu and Zhang [14] or Guo et al. [15]. For example, Liu and Agichtein [13] applied Probabilistic Latent Semantic Analysis (Hofman [16]) to capture user interests in terms of topics based on their answering history. Guo et al [15] developed a probabilistic generative model to obtain latent topics for questions and users.

Nevertheless, current methods of recommending an answerer suffer from limits. As data of users behavior is only collected from CQA services, there are at least three scenarios to which current methods fail to respond:

- User  $A$  has contributed a lot on the collaboration of subject  $S$ , but  $A$  has not yet performed in the CQA of  $S$ .
- User  $A$  has performed well on the subject  $M$ . Subject  $S$  and  $M$  are semantically related.
- User  $A$  has responded to a question on subject  $S$

referring an article which is highly rated in other collaborative activities.

In the above three scenarios, even though  $A$  is most probably an expert on  $S$ , existing methods would not recommend  $A$  as an answerer of  $S$ . We believe that our method to combine analyzing user performance in CQA with that of other collaborative activities could deal with the problem.

## III. OUR APPROACH

In this section, firstly we introduce the scenario and problem we attempt to solve. Then we make a fast review of our previous work about traces in a semantic context. Finally we propose our method of evaluating participant’s performance in CQA service.

### A. Problem statement

This sub-section is divided into two parts: firstly, we explain why CQA and other forms of collaboration need to be combined. Secondly, we introduce show how users are classified by how active they are during collaboration in the community of practice. Users can be recommended according to how active they are.

We found an example from Quora in which six users voted for the answer to the question “What are some examples of well-written academic papers”. We notice that for most scenarios when some answerer needs more detailed evidence or documents to cite to help proving his point, he has to quote external resources by a web-link. This mainly causes two problems: firstly participants temporally step out of the CQA, which leads to the interruption of consistency of reading and furthermore could cause a loss of fidelity to the community; secondly, participants traces of access to and interest in these resources promote the value of resources and in return help improve the reliability of the answer. But as the resource are cited from outside the CQA service, it is impossible to collect traces about the resource such as the quality of the resource or the author of this resource. Collecting that information however, can ameliorate recommending an answerer.

We propose to combine CQA services with other collaborative tools so that firstly when the asker needs to cite a document to demonstrate his idea, he refers to resources he previously stored or read within but not outside of the collaborative platform. At the same time, any users in this platform that have created or read this resource could be potential candidates to answering this question.

### B. Our Proposal

In this section we discuss the features that we take into account for evaluating the performance of users in CQA. In addition, we propose to use the Logistic Regression (Cox [17]) to measure them.

1) *Previous Work on Semantic Traces*: In our previous work, the model of traces proposed by Li [18] allowed an elaborate analysis of interactions among users. We also proposed a model of competency (Wang et al. [19]). In Figure 2, competency is supported by “Interaction activity” in a digital platform including creating, adding, modifying, sharing,

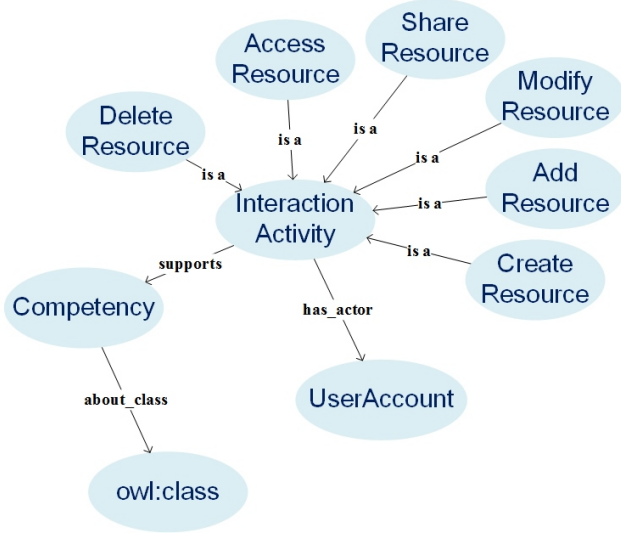


Fig. 2. Generic model of competency [19].

TABLE I. FEATURES BY ROLES OF ASKER AND ANSWERER.[2]

| Features                             | Description   |
|--------------------------------------|---|
| Asker User Traces                    |   |
| Answer/Question Ratio                | Ratio of # of answers to # of questions   |
| Self-voted best answer Ratio         | Ratio of # of best answers that asker himself also voted to # of best answers in the questions posted |
| Total Questions Posted               | # of questions proposed in the past   |
| Answerer User Traces                 |   |
| Best Answer Ratio                    | Ratio of # of best answers to # of total answered questions   |
| Percentage of Vote                   | Percentage of users who voted his answers as the best   |
| Average of Vote for past questions   | Average # of votes for each answer he received for all questions responded                            |
| Average Vote by Asker                | Average # of vote of this answers by the questions' askers  |
| Average Words per Answer             | # of characters in the answer   |
| <i>Remark: # is short for number</i> |   |

accessing and deleting a resource. Each interaction activity is realized through a “UserAccount” representing a user of the platform. “Competency” is also related to a specific subject, e.g., the competency on “Java”.

2) *Features in CQA*: On each semantic concept and for each user, their performance of CQA can be rated by features indicated below. Features are divided into two categories: Asker User Traces and Answerer User Traces. Features in the category of Asker User Traces evaluate a user when he acts as the role of asker including average number of answers for his questions (which indicates if his view of this field is interesting others or not), the ratio of questions whether he himself voted for the best answer (which indicates a consensus of this field), etc. Features in the category of Answerer User Traces evaluate a user when he acts as an answerer including the best answer ratio (the ratio that his answer is voted as the best answer), etc. The complete list of features we consider are listed in Table I.

3) *Logistic Regression*: Our approach consists in training a regressor over seven features from Table I along with traces were left on the platform. Comparing with other regression algorithms that could be applied for our purpose, Logistic Regression [17] is adaptable to our case due to its high variability and non-linear distribution of a variety of input

features. Whatever the input  $t$  is, output  $H(t)$  is always restricted to a rational set  $(0, 1)$ . It uses the logistic function to model an output variable:

$$H(t) = \exp(t)/(1 + \exp(t)) \quad (1)$$

A set of examples is presented by  $(x_i, y_i) \in R^n \times B$  for the training set with  $n$  the number of features. In the training set, for each  $x_i$  the corresponding  $y_i$  is equal to 1 or 0 indicating whether  $x_i$  belongs to a certain class or not. The parameter vector of model  $w \in R^n$  determines the weight of each dimension of vector  $x_i$ . We define the vector of features of user  $i$  on concept  $j$  as  $x_{i,j} = [x_{i,j,(1)}, \dots, x_{i,j,(n)}]$ . Each of the  $n$  dimensions corresponds to features including but not limited to Table I and other features concerning use’s performance in the platform. We will explain the detail of features in Section IV. Therefore (1) becomes:

$$H_w(x_{i,j}) = \exp(w^T x_{i,j})/(1 + \exp(w^T x_{i,j})) \quad (2)$$

where  $H_w(x_{i,j})$  equals to the probability if  $y_{i,j} = 1$  given  $x_{i,j}$  and  $w$ .

$$H_w(x_{i,j}) = P(y_{i,j} = 1 | x_{i,j}, w) \quad (3)$$

#### IV. FIRST PROTOTYPE

In this section, we propose a prototype on MEMORAE (Abel et al. [20]), a web based collaboration platform. Firstly, we introduce the semantic model of this platform. Then, using a toy example we demonstrate how the recommender system is applied.

##### A. Prototype of MEMORAE with CQA

MEMORAE is a web-based platform developed using web 2.0 technologies. The aim is to facilitate knowledge sharing within organizations as discussed by Atrash et al. [21]. In this platform, different types of knowledge resources are supported: social (e.g. chat, event, wiki) and documentary. All types of knowledge resources (e.g. notes, documents, forums, etc.) are indexed by at least one concept of a semantic map. We incorporate CQA services in this platform so that when a question is asked by a user, it is indexed by a concept in the map. Just like the other types of resources: documents, notes, wikis, etc., each answer to a question is indexed according to the question. Figure 3 demonstrates a part of the semantic model and an instance of answering a semantic related question on MEMORAE. A user “P1” uses a user account “UA1” to participate the CQA in the sharing space “S1” of which she is a member. She answers question by an activity “ACR1” indexed by an index key “IK1”. This question is about Java. So the activity “ACR1” supports judging the competency of “P1” on Java. With this trace and other features collected from the platform, we could evaluate competency of “P1” on “Python” as “Java” and “Python” both semantically belongs to the class “ProgrammingLanguage”.

We also deployed the function of voting so that users evaluate a resource by relevance of which it is indexed and sharing space where the resource is visible. As shown in Figure 4, vote is a simpleResource in the ontology of MEMORAE. It is related with IndexKey by the relation hasTarget as shown in Figure 5.

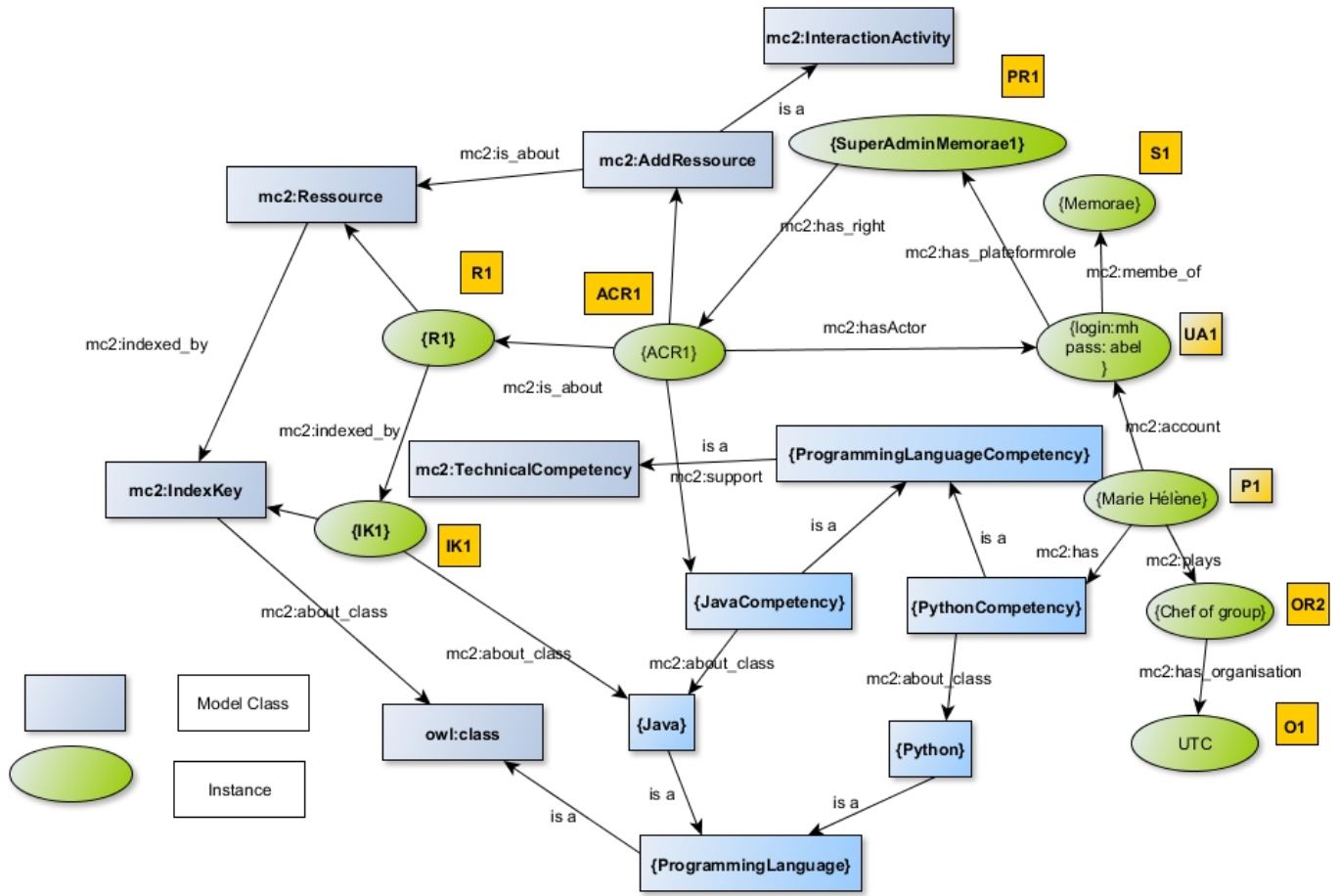


Fig. 3. Semantic model and an instance of answering a semantic related question on MEMORAE.

```

4693 <!-- http://www.hds.utc.fr/mc2/sme/tbox#Vote -->
4694
4695 <owl:Class rdf:about="http://www.hds.utc.fr/mc2/sme/tbox#Vote">
4696   <rdfs:subClassOf rdf:resource="http://www.hds.utc.fr/mc2/sme/tbox#SimpleResource"/>
4697 </owl:Class>

```

Fig. 4. A part of OWL code of MEMORAE.

```

1440 <!-- http://www.hds.utc.fr/mc2/sme/tbox#has_target -->
1441
1442 <owl:ObjectProperty rdf:about="http://www.hds.utc.fr/mc2/sme/tbox#has_target">
1443   <rdfs:range rdf:resource="http://www.hds.utc.fr/mc2/sme/tbox#IndexKey"/>
1444   <rdfs:domain rdf:resource="http://www.hds.utc.fr/mc2/sme/tbox#Vote"/>
1445 </owl:ObjectProperty>

```

Fig. 5. A part of OWL code of MEMORAE ontology showing the relation between vote and other resources.

Vote has four properties: creator, value of vote, IndexKey and date:

- creator: an object property indicating the user who votes.
- value\_vote: a datatype property to describe to what extent a User votes a resource. This value is an integer between 1 and 5.
- IndexKey: an object property that indexes the resource on which people voted, a concept of the ontology this resource is about and a sharing space in which this resource is visible.

TABLE II. AN EXAMPLE OF THREE DIFFERENT INDEXKEYS

| IK1:   | IK2:   | IK3:   |
|--|--|--|
| about_instance: Cpt1;<br>visible_for: S1;<br>index: A1 | about_instance: Cpt2;<br>visible_for: S1;<br>index: A1 | about_instance: Cpt1;<br>visible_for: S2;<br>index: A1 |

- date: a datatype property indicating the date a user votes.

Here for example we show how these properties are defined. Suppose we have an answer A1 to the question Q1. This answer is indexed by concepts Cpt1 and Cpt2 visible in sharing space S1 and also by the concept Cpt1 in sharing space S2. Thus this resource has three Index keys respectively Ik1, Ik2 and Ik3 as shown in Table II. Suppose on 10<sup>th</sup> Oct. 2015, User U1 votes 4 as he believes A1 is very relevant and inspiring about Cpt1 for the members of S1. At the same time User U2 in the same space S1 does not agree and he votes 2 instead. Moreover, U1 votes 2 to A1 about Cpt2 visible in S2 as its much less relevant. Lastly, as S1 and S2 have different members and U1 believe this resource is too deep about concept Cpt1 for the members of S2 to comprehend, so he votes 1 which is visible in S2. After these three activities, we have three resources of vote as shown in Table III:



TABLE III. THREE DIFFERENT VOTES AND THERE CORRESPONDING INFORMATION.

|  |  |  |  |
|--|--|--|--|
| Vote1:<br>creator: U1;<br>value_vote: 4;<br>indexKey: Ik1;<br>date: 10-10-2015 | Vote2:<br>creator: U1;<br>value_vote: 2;<br>indexKey: Ik2;<br>date: 10-10-2015 | Vote3:<br>creator: U1;<br>value_vote: 1;<br>indexKey: Ik3;<br>date: 10-10-2015 | Vote4:<br>creator: U2;<br>value_vote: 2;<br>indexKey: Ik1;<br>date: 10-10-2015 |
|--|--|--|--|

### B. Implementation of Recommender System

Let us have a review of the three scenarios that state-of-art methods could hardly handle: Senario 1

- 1) User  $i$  has contributed a lot on the collaboration of subject  $j$ , but  $i$  has not yet performed in the CQA of  $j$ .
- 2) User  $i$  has responded a question on subject  $j$  referring an article which is highly voted in other collaborative activities.
- 3) User  $i$  is competent on subject  $k$ . Subject  $j$  and  $k$  are semantically related.

For the vector of users features  $X_{i,j}$ , it's far from enough to only include features in Table I. To deal with Scenario 1, performance of user  $i$  on subject  $j$  should be taken into account including all interactions in the platform e.g., how many times  $i$  gets access to resources on  $j$  or how the resources he adds to the platform is voted by members of the sharing space. As resources are evaluated and given different weight, Scenario 2 can be treated. For Scenario 3, we include the probability that  $i$  is expert on  $k$   $H(X_{i,k})$  to evaluate  $H(X_{i,j})$ , i.e.,  $H(X_{i,k}) \in X_{i,j}$ .  $k$  and  $j$  are semantically related.

In Figure 6, we introduce Q&A Performance Set, User Profile and Traces in the platform as the training set  $T(X_{i,j}, Y_{i,j})$  to train the Logistic Regression. To determine  $w$ , we minimize the squared loss by a training set of  $m$  samples:

$$L(w) = \sum_{i=1}^m (H_w(x_{i,j}) - y_{i,j})^2 + \frac{\lambda}{2} \|w\|^2 \quad (4)$$

where  $\frac{\lambda}{2} \|w\|^2$  is applied for regularization to avoid overfitting. With the Logistic Regression model trained, we give recommendation on  $j$  by comparing  $H(X_{x,j})$  where  $x \in \{\text{members of the sharing space}\}$ .

We illustrate the application of Logistic Regression by a fictitious example. Table IV contains a training set including data of seven users and a testing set. In this data set, variables  $x_1, \dots, x_7$  represent figures in Table I. For example,  $x_3$  of user A means total questions posted by user A is 7.  $x_8$  is the number of activities other than CQA on C1 a user participates.  $x_9$  is the total votes received by the resources a user cites in the CQA service.  $x_{10}$  represents the probability of user being a competent answerer on concept C2, which is semantically close to C1. In the training set users are distinguished by labels. A label of 1 means a competent answerer to recommend while 0 is not. We substitute this training set into equation 4 and minimize the loss  $L(w)$  to get a solution of parameter set  $w$ . Back to equation 2, with the acquired  $w$  and test set in Table IV, we obtain  $y_1 = 0.87$ ,  $y_2 = 0.03$ . It means user 1 has a probability of 0.87 to be a competent answerer. On the contrary user 2 is not recommended as the corresponding probability is only 0.03.

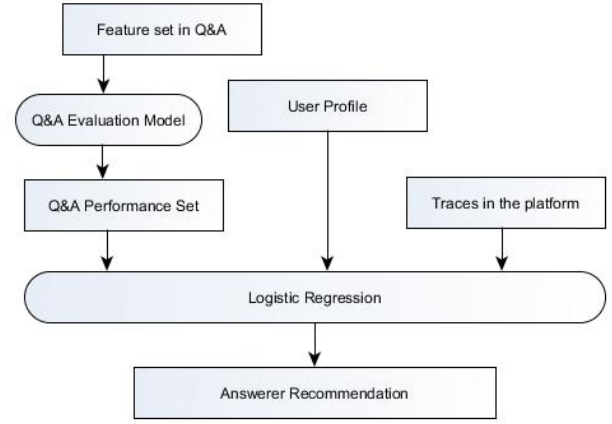


Fig. 6. The approach of recommender system based on features of CQA service and traces in the platform.

From the previous example, one can appreciate that our approach works well in distinguishing users. Three aspects contribute to this result. Firstly, we evaluate users not merely on their interactions on CQA, but on all activities in the collaborative platform. Secondly, how the resource users refer to is evaluated is also considered as a feature in the evaluation. Thirdly, users are not only evaluated by a single concept on which the newly arrived question is asked, but also by the semantically related concepts.

### V. CONCLUSION AND FUTURE WORK

With great power comes great responsibility. This paper presents an approach to evaluate users' competency given traces in the CQA service and other collaborative context. To reach this goal, we propose to combine the analysis of a variety of figures in CQA services with our previous work to measure users performance in a collaborative environment. A gap can be bridged between sharing knowledge resource and CQA service to increase the reliability of an answerers competency.

We integrate a stereotype of semantic CQA service in a collaborative context. From a global view we use our proposed approach to evaluate users by the performance in the CQA service and the corresponding knowledge sharing activities he participated in a collaborative context. After modeling and analyzing the behavior of users, we pair a potential user to answer a newly proposed question.

Further work includes collecting data from students of a course entitled "IA03" and evaluate the performance of our model. We apply the system of collecting users' traces in a real scenario where participants are students enrolled in "IA03", a course on knowledge management in the University of Technology of Compiègne. At the end of this semester we will analyze each participant's competency on each concept as a reference to their final score.

### REFERENCES

- [1] Y. Liu, S. Li, Y. Cao, C.-Y. Lin, D. Han, and Y. Yu, "Understanding and summarizing answers in community-based question answering services," in *Proceedings of the 22nd International Conference on*

TABLE IV. EXAMPLE OF TRAINING SET AND TESTING SETS

| Training set |       |          |       |       |       |       |       |       |       |       |          |
|--------------|-------|----------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| User         | Label | features |       |       |       |       |       |       |       |       |          |
|              |       | $x_1$    | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
| A            | 0     | 3.50     | 0.20  | 7.00  | 0.50  | 0.10  | 4.20  | 55.00 | 0.60  | 2.00  | 0.30     |
| B            | 1     | 8.50     | 0.60  | 5.00  | 0.30  | 0.05  | 5.00  | 70.00 | 0.40  | 8.00  | 0.70     |
| C            | 0     | 15.00    | 0.80  | 18.00 | 0.10  | 0.02  | 1.50  | 28.00 | 0.10  | 3.00  | 0.20     |
| D            | 1     | 9.50     | 0.30  | 4.00  | 0.90  | 0.15  | 5.00  | 56.00 | 0.80  | 12.00 | 0.78     |
| E            | 1     | 2.50     | 0.60  | 20.00 | 0.70  | 0.25  | 4.00  | 5.00  | 0.60  | 15.00 | 0.68     |
| F            | 0     | 5.00     | 0.40  | 38.00 | 0.10  | 0.02  | 1.50  | 46.00 | 0.17  | 3.00  | 0.21     |
| G            | 0     | 4.00     | 0.30  | 48.00 | 0.20  | 0.12  | 1.80  | 56.00 | 0.27  | 6.00  | 0.22     |
| Testing set  |       |          |       |       |       |       |       |       |       |       |          |
| 1            | $y_1$ | 3.50     | 0.20  | 12.00 | 0.70  | 0.12  | 2.80  | 0.00  | 0.20  | 6.00  | 0.30     |
| 2            | $y_2$ | 2.50     | 0.60  | 20.00 | 0.70  | 0.12  | 1.80  | 5.00  | 0.80  | 8.00  | 0.22     |

*Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008, pp. 497–504.

- [2] Q. Liu, E. Agichtein, G. Dror, E. Gabrilovich, Y. Maarek, D. Pelleg, and I. Szpektor, “Predicting web searcher satisfaction with existing community-based answers,” in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 415–424.
- [3] D. R. Raban, “Self-presentation and the value of information in q&a websites,” *Journal of the American society for information science and technology*, vol. 60, no. 12, pp. 2465–2473, 2009.
- [4] S. A. Paul, L. Hong, and E. H. Chi, “Who is authoritative? understanding reputation mechanisms in quora,” *arXiv preprint arXiv:1204.3724*, 2012.
- [5] E. Negre, *Information and Recommender Systems*. John Wiley & Sons, 2015.
- [6] X. Liu, W. B. Croft, and M. Koll, “Finding experts in community-based question-answering services,” in *Proceedings of the 14th ACM international conference on Information and knowledge management*. ACM, 2005, pp. 315–316.
- [7] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, “Finding high-quality content in social media,” in *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, 2008, pp. 183–194.
- [8] C. Shah and J. Pomerantz, “Evaluating and predicting answer quality in community qa,” in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2010, pp. 411–418.
- [9] J. Lin and D. Demner-Fushman, “Methods for automatically evaluating answers to complex questions,” *Information Retrieval*, vol. 9, no. 5, pp. 565–587, 2006.
- [10] J. Lin and P. Zhang, “Deconstructing nuggets: the stability and reliability of complex question answering evaluation,” in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2007, pp. 327–334.
- [11] P. Jurczyk and E. Agichtein, “Discovering authorities in question answer communities by using link analysis,” in *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, 2007, pp. 919–922.
- [12] M. Bouguessa, B. Dumoulin, and S. Wang, “Identifying authoritative actors in question-answering forums: the case of yahoo! answers,” in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, pp. 866–874.
- [13] Q. Liu and E. Agichtein, “Modeling answerer behavior in collaborative question answering systems,” in *Advances in information retrieval*. Springer, 2011, pp. 67–79.
- [14] B. Liu and L. Zhang, “A survey of opinion mining and sentiment analysis,” in *Mining text data*. Springer, 2012, pp. 415–463.
- [15] J. Guo, S. Xu, S. Bao, and Y. Yu, “Tapping on the potential of q&a community by recommending answer providers,” in *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008, pp. 921–930.
- [16] T. Hofmann, “Probabilistic latent semantic indexing,” in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999, pp. 50–57.
- [17] D. R. Cox, “The regression analysis of binary sequences,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 215–242, 1958.
- [18] Q. Li, “Modeling and exploitation of the traces of interactions in the collaborative working environment,” Ph.D. dissertation, Université de Technologie de Compiègne, 2013.
- [19] N. Wang, M.-H. Abel, J.-P. Barthes, and E. Negre, “Mining user competency from semantic trace,” in *Computer Supported Cooperative Work in Design (CSCWD), 2015 IEEE 19th International Conference on*. IEEE, 2015, pp. 48–53.
- [20] M.-H. Abel, A. Benayache, D. Lenne, and C. Moulin, “E-memorae: a content-oriented environment for e-learning,” in *E-Learning Networked Environments and Architectures*. Springer, 2007, pp. 186–205.
- [21] A. Atrash, M.-H. Abel, and C. Moulin, “Notes and annotations as information resources in a social networking platform,” *Computers in Human Behavior*, 2015.