International Journal of Natural and Engineering Sciences 7 (2): 82-85, 2013
ISSN: 1307-1149, E-ISSN: 2146-0086, www.nobel.qen.tr

A novel method based on concept map for expert finding in online communities

Ahmad KARDAN[1] Majid RAFIEI[1*]

[1]Advanced E-Learning Lab, Dep. of Computer Eng. and IT, Amirkabir University of Technology, Tehran, IRAN

*Corresponding author:
E-mail: mrafiei@aut.ac.ir

Received: January 04, 2013
Accepted: February 17, 2013

Abstract

Nowadays, online communities or virtual communities on the Internet are one of the most interactive environment in which people can express their opinions freely. There is too amount of information shared in online communities and there is no way to determine its authenticity, this is the most important challenge in the field of knowledge sharing in online communities. Expert finding in the online communities and determine their level of knowledge, can be used to determine the accuracy of posted comments. In this paper a new method for expert finding in online communities based on concept mapping, is presented. In the proposed method, two measures are used. One measure is distance between concepts of user response and concepts of question. Other measure is the number of concepts that are used in the user response. Proposed method is implemented and evaluated on Java Online Communities, and the results showed that the correlation exceeds 0.9.

Keywords: Expert finding, Online community, Concept map, Dijkstra's algorithm, Knowledge sharing

INTRODUCTION

Nowadays, people from all over the world can freely share their questions or comments in online communities and or discuss about other users comments and posts. Online communities are one of the important achievements of Web2.0 technologies, and due to the large amount of information, from academic researchers and commercial organizations in recent years have been welcomed [1].

One of the most important applications of online communities in virtual space of the Internet is knowledge sharing. Online communities, due to its unique features such as ease of access and lack of time/place constraints, have been able to become one of the main sources of problems solving.

Virtual communities can be divided into two general categories: professional virtual communities and public virtual communities. In professional virtual communities usually technical discussions relating to a specific subject are done, while public subjects are discussed in public virtual communities. Some of the reasons of participation in professional virtual communities includes: problem solving, improve the individual's ability, attract specific knowledge and innovation [2].

The lack of distinction between users level of knowledge and unknown value of answers and comments, are biggest challenge in online communities. In fact, one that ask question, does not know, how much confidences to the answer submitted by other users. With expert finding and determine the users level of knowledge, we can determine what answers is more reliable.

Another important challenge in online communities is large amount of information relates to questions posted by users. It makes question has little chance of being seen by experts that have the ability to respond to the question and therefore one that ask question, does not get the correct answer or response time is longer. With techniques of expert finding and using Recommender Systems, questions can be exposed to individuals who have abilities and knowledge that are necessary to respond. Also the simple questions that is not difficult, do not display to experts to they spend their time just to answer the questions that people are incapable of responding to it. This Recommender System has been implemented in [3].

Also with expert finding can summarize large amount of information in online communities. So that answers submitted by users are validated according to the sender level of knowledge and only answers are displayed that the sender had the knowledge necessary to answer. Thus, those who

seek to find answers in online communities not be confused with large amount of information.

Given the above, it is clear that the methods of expert finding in online communities is very important in order to better exploit the mass of valuable information contained in these communities. In this paper, a new method for expert finding in online communities is presented. Section 2 introduces the related works has been done in this field. What needed for a better understanding of the presented method, are expressed in section 3. The proposed method is presented in section 4 and in section 5, the used methodology is explained. In section 6, the proposed method is evaluated and finally the conclusion of the study is described.

RELATED WORKS

The problem of finding experts has been one of the most top issues for researchers since more than 15 years ago. In the past, most research in the field of expert finding had been done in organizations, but now most of the researches are done on the Internet, especially in social networks and online communities [4]. Some of the works that are described in this section are related to the organization and some others are related to the Internet.

Expert finder systems are part of the CSCW (Computer Supported Cooperative Work) systems that help to find people with expertise or special information. For example, to find people in an online community that have the ability to respond to a particular question. Expert finder systems are an important class of the recommender systems [3].

In general, there are two main approaches to finding experts. The first approach focuses on analyzing social networks. For instance, Network-based ranking methods including PageRank and HITS have been used in order to identify experts. The basic idea of network-based ranking algorithms is that people in social networks are considered as nodes and relationship between them are considered as edges in a graph. Forming an edge between the two sides are required to exchange information between them. For example, if person A respond to person B in a discussion group, an edge from node that indicates A is drawn to node that indicates B. After creating of such a network is also called the Expertise Network [5], network-based ranking algorithms are able to find important nodes in the network that indicates the experts. For example, in [5] aim is to find different ways to identify and rank the experts in a field based on formation Expertise Network and compare the performance of these methods. In [4] the aim is to finding experts in MetaFilter online community, using social network analysis approach. [6] Finds experts with social network analysis approach in Friendfeed online community.

The second approach to finding experts in an online community focused on content analysis. In this approach using text mining techniques, the messages that sent by users are analyzed and based on information extracted from a text messages a user knowledge model using user-knowledge modeling or a probability model of the relationship between the user and the messages is generated. With using knowledge model or probability model, expert users can be identified. For example in [1] user knowledge modeling has been used to identify experts. In [7], [8], [9] and [10] experts have been identified using a probabilistic model.

BASIC CONCEPTS

This section presents the concepts needed for a better understanding of the proposed method that contains three subsections, includes: what is Java online community and how it works, what the Concept Map is, what Dijkstra's algorithm is and how it works.

Java Online Community

Java online community is part of Oracle corporation forum that is related to the Java technology. According to information we obtained in this study, until February 2013 Oracle forum has almost a million users and almost two million and a half questions in the forums has been raised and examined. These statistics clearly indicates that this online community is very active.

Join the Online Community membership is free, and users can post their questions on the forum after membership. Previous FAQ threads view is not required to registration and there are visible to everyone. This forum has 16 subsections, each corresponding to one of the following Java technologies. This segmentation has led to questions posted in this forum were highly specialized.

Like most online communities online community Java has a scoring mechanism to the users, so that the inquirer user can use two types of labels for respond that submitted by other users. If submitted answer get "Helpful" label by the inquirer user, respondent user receives 5 points, and if submitted answer get "Correct" label by the inquirer user, respondent user receives 10 points. The points in each subsection of the online community are gathered and so the rating of each user in the mentioned subsection is identified. 10 expert users in each subsection are defined based on these points. At the end, total scores in all the subsections are collected and overall scores are determined. In total 10 expert users in online community according to these points are introduced, regardless of the section in which they operate. In this study, we evaluate our proposed method by calculate correlation between scores of our method and 10 expert user in each subsection.

Concept Map

Concept map is a graphical way to represent knowledge, this map is actually a graph that contains nodes to represent concepts and labeled links explain the relationship between the concepts [11]. Concept mapping can also be used as a method to infer students' conceptual knowledge in a particular field [12].

In this study, Java technology concept map has been used as the basis for extracting the concepts used in the question and answer, and also for calculating distance between concepts of user response and concepts of question.

Dijkstra's algorithm

Dijkstra's algorithm, conceived by computer scientist Edger Dijkstra in 1956 and published in 1959 [13], Dijkstra's algorithm is a graph search algorithm that solves the single-source shortest path problem for a graph with non-negative edge path costs, producing a shortest path tree.

For a given source vertex (node) in the graph, the algorithm finds the path with lowest cost (i.e. the shortest path) between that vertex and every other vertex. It can also be used for finding costs of shortest paths from a single vertex to a single destination vertex by stopping the algorithm once the shortest path to the destination vertex has been determined. For example, if the vertices of the graph represent cities and edge path costs represent driving distances between pairs of cities connected by a direct road, Dijkstra's algorithm can be used to find the shortest route between one city and all other cities.

In this paper, we use Dijkstra's algorithm for finding the shortest path between concepts in concept map. Indeed, the concept map is a weighted graph where all edge weights are equal to one.

PROPOSED METHOD

As mentioned for estimating the user's knowledge, a concept map about Java technology has been used. At the beginning, posts that sent by the user are extracted, in these posts, subject of question, question text and answer text submitted by the user is available. After extracting this information, concepts used in the text of the question and also concepts used in the text of the answer, according to existing concepts in the concept map are extracted.

Our proposed method based on two measures get scores to users, which includes:

- Distance between concepts of user response and concepts of question.
- The number of concepts that are used in the user response

In order to calculate distance between concepts of user response and concepts of question, we need to get shortest distance between two concepts in the concept map. To do this, the relationship between two concepts in the concept map is extracted, and the graph is formed from these concepts, and relationship between them. With using Dijkstra's algorithm, the shortest path between the two concepts are extracted and stored in a matrix.

Distance between concepts of user response and concepts of question, can be calculated in three ways, as follows:

1. Extracting the shortest path between concept of user response and each one of the concepts in the text of question and averaging them.
2. Extracting the shortest path between concept of user response and each one of the concepts in the text of question and choose their minimum.
3. Extracting the shortest path between concept of user response and each one of the concepts in the text of question and choose their maximum.

Here we use the first way. After obtaining the number of concepts that are used in the user response and distance between concepts of user response and concepts of question, user's score is calculated using formula 1.

$$Score(P_i) = \sum_{M \in Messages} \left(\sum_{R \in Responses} \alpha . Rep(C_{MR}) + \frac{\beta}{\sum_{Q \in Questions} (Dist(C_{MR}, C_{MQ})) / N_{MQ}} \right) \quad (1)$$

$R,Q \in$ Concepts of Concept Map

In formula 1:

- Score(P_i): Scores of user's *i*
- Messages: Messages of user's *i*
- Responses: Concepts in the response of the message *i*
- Questions: Concepts in the question of the message *i*
- Rep(C_{MR}): The number of iterations of concept R in the response of the message *M*
- Dist(C_{MR}, C_{MQ}): Distance between concept *R* in the response of the message *M*, and concept *Q* in the question of the message *M*
- N_{MQ}: The number of concepts in the question of the message *M*

In formula 1, $\sum_{Q \in Questions} (Dist(C_{MR}, C_{MQ})) / N_{MQ}$, calculates the average of the distance between concept *R* in the response of the message *M*, and all the concepts *Q* in the question of the message *M*. This value is calculated for all of concepts in the response of the message *M*. In the numerator, sum of the distances of concept *R* in the response of the message M from each of the concepts *Q* in the question of the message *M*, are calculated. And in the denominator, N_{MQ} is the number of existing concepts in the question of the message M.

α and β are coefficients with values between 0 and 1, respectively indicates the impact of the number of concepts that are used in the user response, and the impact of distance between concepts of user response and concepts of question. Here the optimal values for these coefficients are calculated, so using these coefficients, the best correlation between the scores of the proposed method and the scores by Java's online community, will be obtained. Values that have been obtained for these coefficients are 0.5 for both of them.

EVALUATION AND RESULTS

To evaluate the proposed method, all the subsection of Java online community is used. First, number of responses for each subsection was calculated and subsections that the number of responses for them is less than 3000 have been excluded. Finally, 11 subsections have remained.

Spearman correlations were calculated separately for the 11 subsection and the overall correlation is calculated by taking the average of the subsection correlation. Correlation was also calculated for the entire Java forum. Mean Spearman correlation was obtained equal to 0.902 for all these cases.

Table 1 shows information for each subsection, separately. In this table, the meaning of abbreviations is:

- *NQ*: Number of Question
- *NR*: Number of Response
- *NU*: Number of active Users
- *SpCo*: Spearman correlation

Table 1. Information about each subsection

Category	NQ	NR	NU	SpCo
ALL	6465	345206	614	***0.88***
Database Connectivity	367	23456	254	***0.84***
Development Tools	308	4869	152	***0.94***
Java APIs	337	31096	105	***0.93***
Java Card	415	4145	28	***1***
Java Desktop	1657	54839	150	***0.94***
Java Developer Tool APIs	9	2719	37	**-**
Java Embedded	8	62	13	**-**
Java Enterprise & Remote Computing	447	40642	134	***0.82***
Java Essentials	2266	157398	264	***0.91***
Java HotSpot Virtual Machine	40	8299	51	***0.86***
Java Mobile	80	2000	30	**-**
Java Real-Time	0	12	4	**-**
Java Security	86	4868	50	***0.85***
Java TV	0	5	3	**-**
Java FX	484	9689	53	***0.94***
Other Topics	58	5905	47	***0.87***

For 'Java Developer Tool APIs', 'Java Embedded', 'Java Mobile', 'Java Real-Time', 'Java TV', because the number of responses is less than 3000, the correlation is not calculated for them.

In table 1, the value of NR is more important for our study, because the proposed method is based on user's responses and if these numbers are much higher, accuracy will be higher.

CONCLUSIONS AND FUTURE WORKS

In this study, the role and importance of online communities was discussed. In this regard, in addition to understanding knowledge sharing and its position in online communities, important concerns and challenges were expressed in online communities, and on one of the solutions to these challenges were focused that was "Expert Finding", related works done in the field of "Expert Finding" were expressed, and a new method based on concept maps for expert finding in online communities were presented. In this method, two measures were used for estimating users' knowledge. One measure was distance between concepts of user response and concepts of question. Other measure was the number of concepts that are used in the user response. The proposed method is implemented and evaluated on Java Online Communities, and the results showed that the correlation exceeds 0.9.

The proposed method in this study, only works based on content analysis, and networking among users has not incorporated. In the future, we can add the influence of communication between users based on social network analysis methods, and reach to better results. For this purpose, link analysis algorithms such as PageRank or HITS, can be used as an effective treatment.

REFERENCES

[1] A. Kardan, M. Garakani, and B. Bahrani, "A method to automatically construct a user knowledge model in a forum environment", SIGIR '10 Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, 2010.

[2] Chih-Jou Chen, Shiu-Wan Hung, "To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities", Journal of Information & Management, Elsevier, 4 (May 2010), 226-236.

[3] J. Zhang et al, "QuME: A Mechanism to Support Expertise Finding In Online Help-seeking Communities", 20th annual ACM Symposium on User Interface Software and Technology, 2007, USA, 111-114.

[4] A. A. Kardan, A. Omidvar, M. Behzadi, "Context based Expert Finding in Online Communities using Social Network Analysis", International Journal of Computer Science Research and Application, Vol. 02, Issue. 01(Special Issue), 2012, pp. 79-88.

[5] J. Zhang et al, "Expertise Networks in Online Communities: Structure and Algorithms", Proceedings of the 16th International World Wide Web Conference (WWW2007), Banff, Canada, 2007, USA, 221-230.

[6] A. Kardan,A. Omidvar, F. Farahmandnia, "Expert Finding on Social Network with Link Analysis Approach", Electrical Engineering (ICEE), 19th Iranian Conference, 2011.

[7] X. Liu, W. B. Croft, M. Koll, "Finding experts in community-based question-answering services", Proceedings of the 14th ACM international conference on Information and knowledge management, Germany, 2005.

[8] P. Serdyukov, H. Rode, D. Hiemstra, "Exploiting sequential dependencies for expert finding", Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, Singapore, 2008.

[9] K. Balog, L. Azzopardi, M. Rijke, "Formal models for expert finding in enterprise corpora", Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, USA, 2006.

[10] K. Balog and M. Rijke, "Determining Expert Profiles (With an Application to Expert Finding)", IJCAI'07 Proceedings of the 20th international joint conference on artifical intelligence, 2007.

[11] Novak, J. D. and Canas, A. J. 2008. "The Theory Underlying Concept Maps and How to Construct and Use Them", Technical Report IHMC CmapTools 2006-01 Rev 01-2008.

[12] Park, U., Calvo, RA. "Automatic Concept Map Scoring Framework Using the Semantic Web Technologies", Proceedings of Eighth IEEE International Conference on Advanced Learning Technologies, Santander, Cantabria, 2008.

[13] Dijkstra, E. W. "A note on two problems in conanexion with graphs", Numerische Mathematik 1:269–271, 1959.