A Point-Based Semi-Automatic Expertise Classification (PBaSE) Method for Knowledge Management of an Online Special Interest Group

Aisyah Ismail,
Shahida Sulaiman, Maziani Sabudin
School of Computer Sciences,
Universiti Sains Malaysia,
11800 USM, Penang,
Malaysia.
aisyahismail@gmail.com,
shahida@cs.usm.my, maziani@cs.usm.my

Sarina Sulaiman
Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia,
81310, Skudai, Johor,
Malaysia.
sarina@utm.my

Abstract

An online Special Interest Group is a group of people with the same interest gather to form an online community through the Internet. In certain cases where the knowledge is being manipulated, the portal of a Special Interest Group is in a form of knowledge portal. This portal allows interaction among its community members. Interaction through forum in the portal makes activities such as discussion of problems and knowledge sharing among each other possible. The need to classify users' expertise in a Special Interest Group is crucial task. This paper describes a Point-based Semi-automatic Expertise classification method to classify users' expertise in a Special Interest Group knowledge portal.

1. Introduction

Knowledge is information and skills acquired through experience or education. We live in the knowledge era where knowledge is available almost everywhere in abundance. Therefore, knowledge should not be neglected; it needs to be shared and exchanged. Based on Newman and Conrad [1], knowledge management is a discipline that seeks to improve the performance of individuals and organizations by maintaining and leveraging the present and future value of knowledge assets. Knowledge portal is an enhancement of the ordinary web portal. While the web portal focuses on offering users a broad array of resources and services, the knowledge portal does not only offer the resources and services, it also acts as a knowledge repository where it will extract and analyze knowledge submitted among its community members. According to Niwa [2], a knowledge sharing paradigm perceives knowledge supplier as the same set of system users who use the knowledge base. Knowledge portal is one of the means for knowledge sharing.

Based on Giarratano and Riley [3], there are three ways to represent knowledge. They are rules, frames and semantic nets. Rules are the most common type of knowledge representation. Rules are easy to implement due to its straightforward structure. However, ordering of the rules is important. Frames represent related knowledge about an object. Frames are easy to understand and frames allow unrestrained alteration or cancellation of slots. Frames are suitable to describe a mechanical device. Semantic nets are simple, economical and relatively intuitive representation form. Besides, semantic nets are easy to be implemented and manipulated due to its flexibility to cluster related knowledge. The structure of semantic nets is denoted by nodes and arcs. This paper will use semantic nets to represent its knowledge.

In a Special Interest Group (SIG) knowledge portal, people from various backgrounds gather for several reasons. For instance, students join a SIG to derive some guidance from people who are already in the industry. They can also be experts in certain fields who are willing to answer questions from anyone and share their expertise. On the other hand, there are also some people who join the portal simply to make new friends who have the same interest. Likewise, these people posses knowledge and they are willing and eager to share their knowledge with each other through this online community.

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

Having people with various backgrounds in the community, we find the need to classify the users' expertise. Knowledge can be organized by classifying expertise of the user. In other words, users' expertise is the knowledge in the portal. When users join the portal for the first time, they may want to find other users' who share the same interests and problems. In addition, they look for someone in the portal who is an expert in a certain field to seek their help in the problems that they face. Classification of the users' expertise is a very crucial task. Hence, it needs to be handled by the developer of the SIG knowledge portal to ensure the convenience of the community members.

In Section 2 we discuss the related work and the problems that motivate the work. Section 3 describes the proposed method, followed by Section 4 explains the implementation of the proposed solution. Section 5 explains the qualitative evaluation of the proposed method. Finally we conclude our work in Section 6.

2. The motivation

Online communities are not much different from other real world communities. Both communities consist of people who are tied together by their interests. In an online community, a group of people from different backgrounds are strangers to each other and this makes them become keen to get some information about the people in their community. Knowing one's level of expertise will make knowledge sharing and discussion more meaningful. The portal will state users' level of expertise for all community members to view.

This section will discuss the existing classification method in exiting Web portal and the related work in classifying expertise in a SIG portal.

2.1 Existing SIG portals

We study ITTutor.net [4] and Computer Forum [5] to better understand the existing web portal classification method. Both of these web portals are well respected web portal where its registered users reached more than 4000 and the members are increasing.

Both ITTutor.net [4] and Computer Forum [5] rank users based on the number of posts they made in the portal. The more forums posted the higher users' rank will be. By doing so, even when users post query on a certain topic or post something irrelevant to the topic, users' rank will increase. Given a scenario where *A*, who is a total beginner, posts a lot of queries in the forum without really contributing something. Then there is *B*, who on contrary answers other users query in the forum. However, *A*'s posts are larger in number than *B*'s posts. Based on the existing ranking approach, *A* will be ranked higher than *B*, which is inappropriate.

In ITTutor.net [4], there are three ways to identify users' position in the portal (See Figure 1). They are users' status, military-based ranks and rating from other users in the portal. Users' status will be assigned *Core*, *Ahli Biasa* (Normal Member), *Pengendali* (Administrator), *Ahli Professional* (Professional Member) or *Ahli* (Member). However, this users' status is not used to classify the users' expertise.

Figure 1. Three ways to identify users' position in ITTutor.net [4]

Military-based ranks as listed in Table 1 are used to rank the users in the portal. Users are ranked based on points they collected in the portal. The rating function will collect points given by other users in the portal and will be presented as stars. These three ways to identify users' position in the portal will lead to users' confusion of the actual users' level of expertise.

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

Table 1. Military-based rank used in ITTutor.net [4]

Rank	Minimum Points
Kadet	0
Korporal	50
Sarjan	100
Staf Sarjan	150
Sarjan Mejar	200
Pegawai Waran 1	300
Pegawai Waran 2	400
Leftenan Muda	500
Leftenan	1000
Kapten	1500
Mejar	2500
Leftenan Kolonel	3000
Kolonel	3500
Certified ITTutor Professional	10000

The Computer Forum [5] ranks its users based on the minimum posts made by users in the portal as listed in Table 2 below. The way ranks are given is based on the minimum posts made by users in the portal. Other than that, there are also special ranks given by the administrator of the portal to selected users. Administrator also has the rights to ban users who violate the rules and regulations of the portal.

Table 2. Ranks in Computer Forum [5]

Rank	Minimum Posts
New Member	0
Bronze Member	25
Silver Member	100
Gold Member	250
Platinum Member	500
Diamond Member	1000
Unspecified by computerforum.com	2000
Unspecified by computerforum.com	4000
Unspecified by computerforum.com	6000
Unspecified by computerforum.com	8000
Unspecified by computerforum.com	10000

2.2 Expertise classification methods

The existing methods include that of Zhang *et al.* [6] who proposed z-score measures ExpertiseRank that was based on PageRank algorithm proposed by Page *et al.* [7]. In the work of Zhang *et al.* [6] the proposed algorithms were compared with PageRank [7] and HITS (Hypertext Induced Topic Selection) of Kleinberg [8] in a Java forum of an e-community to analyze the relative expertise of different users. The evaluation showed that both ExpertiseRank and z-score performed the best in e-community with different characteristics.

The z-score measures [6] combine both the asking and replying patterns. For example if users ask about the same number of queries and answers, the z-score will be close to 0. If they answer more than asking questions, the z-score will be positive otherwise it will be negative. In addition, ExpertiseRank [6] increases expertise scores using question-answer network. For instance if *A* is able to answer *B*'s questions, and *C* is able to answer *B*'s questions, then *C*'s expertise rank should be promoted because *C* is able to answer *B*'s question where *B* also happen to be someone who has some expertise. Nevertheless, the measures produced are still questionable, as the quality of the answers is not considered in the measures.

On the other hand, HITS [8] rate e-community users based on their authority and hub values in the community network nodes. Authority value is the sum of the scaled hubs values that point to the user and hub value is the sum of the scaled authority values of the user. Users with the highest authority score are experts in the community whilst users with the highest hub values are beginners who have good contact with the experts. Yet the setting of values for authority and hub could be affected if the actual contents of network nodes are of low quality that cause the increased number of authority and hub values when more unnecessary communication occurs.

Another work by Löser and Tempich [9] suggested three semantic overlay layers to give scores to e-community peers using peer monitor based on the frequency to answer a query either as responses to information requests, asking similar questions, providing related documents and asking questions of diverse topics in the past. Peer monitor is a good way that needs users' intervention to rank the peers. However the peers may give unjustified scores that cause discrepancies in the peer monitor.

Hence, this paper proposes a point-based semi-automatic expertise classification that employs z-score of Zhang *et al.* [2] and mapped to a 5-scale point with the combination of a manual classification towards the answers given by the members of a SIG e-community.

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

3. Point-Based Semi-Automatic Expertise (PBaSE) classification method

The proposed classification is called Point-Based Semi-Automatic Expertise (PBaSE) classification method. This is a two-way classification method where the portal will automatically classify users' expertise level based on users' interaction in the portal and users' rating. PBaSE method consists of two parts; automatic classification using z-score measures of Zhang *et al*. [6] and users' rating. PBaSE method takes the average of the two parts as the users' level of expertise. Users are classified as beginner, intermediate and expert based on the accumulated points.

There are two types of post in the portal. They are 'query' post and 'answer' post. The 'query' post is made by a user to ask questions under a certain topic. On the other hand, 'answer' post is a post that answers questions to the 'query' post. Logically, users who make more 'answer' post are the expert and users who make more 'query' post are the beginner in the portal. This paper will be using the z-score measures as introduced by Zhang *et al*. [6] to classify users in the portal.

$$\sum_{i=1}^{n} Z_i = \frac{a_i - q_i}{\sqrt{a_i + q_i}}$$

Let Z_i be the z-score for user i, i =1 until n where n is the number of users, a is the number of 'answer' post made by user and q is the number of 'query' post made by user.

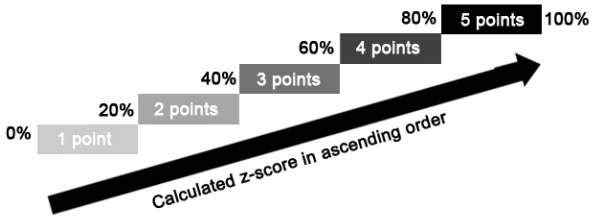

Figure 2. Mapping of the z-score measures

Once the z-score for all users are calculated, the value will be sorted in ascending order and will be mapped to an appropriate point as illustrated in Figure 2. The top 20% of the users will be given 5 points. The last 20% of the users will be given 1 point. The other users will be given points as shown in Figure 2. The top 20% of the users (based on an ascending order of calculated z-score) are the top contributors in the portal and will be given 5 points each. The rationale behind this mapping system is that the experts are always the top contributors of the portal. This means, even when a user is an expert but if the user stops contributing to the portal, the user's level of expertise may drop if there are other users who contribute more. If there is a tie for the highest contributor, both users will be given 5 points.

Table 3 shows an example of mapping the z-score measures. Let U_i be the users, i = 1 until n where n is the number of users, q is the number of queries posted, a is the number of answers posted, Z is the z-score measures [6] and M is the mapped z-score.

When users view 'answer' posts in the portal, they are required to rate by the scales: 0 (extremely poor), 1 (very poor), 2 (poor), 3 (good), 4 (very good) or 5 (excellent). By doing so, the sender of the post will receive points given by the other users. We treat all post equally, thus the user rating points, R is calculated by dividing the total points collected for each user, T with the numbers of users who make the rating, N. The purpose of user rating, R is to counter check the automatic classification using z-score measure [6].

Table 3. An example of measures

				Z-score in ascending order		
U_i	q	a	Z	U_i	Z	M
U_1	5	0	-2.24	U_6	-7.07	1
U_2	0	5	2.24	U_9	-4.08	1
U_3	5	5	0	U_1	-2.24	2
U_4	10	5	-1.29	U_4	-1.29	2
U_5	5	10	1.29	U_3	0	3
U_6	50	0	-7.07	U_8	0	3
U_7	0	50	7.07	U_5	1.29	4
U_8	50	50	0	U_2	2.24	4
U_9	100	50	-4.08	U_{10}	4.08	5
U_{10}	50	100	4.08	U_7	7.07	5

$$\sum_{i=1}^{n} R_i = \frac{T_i}{N_i}$$

The final points (for each user), F is the average of the sum of mapped z-score, M and users' rating, R. The mapping of the final points, F to the expertise level, L are: expert E (4 or 5 points), intermediate I (2 or 3 points) and beginner B (0 or 1 points).

$$\sum_{i=1}^{n} F_i = \frac{M_i + R_i}{2}$$

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

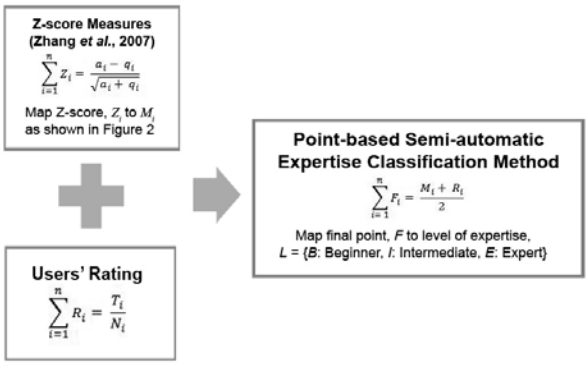

Figure 3. Measures by PBaSE method

Figure 3 illustrates an overview of PBaSE. Let U_i be the users, $i = 1$ until n where n is the total number of users, q is the number of queries posted, a is the number of answers posted, Z is the z-score measures [6], M is the mapped z-score, R is the users' rating, F is the final points and L is the level of expertise {B: Beginner, I: Intermediate, E: Expert}.

An example of classification using PBaSE is as shown in Table 4. Based on the classification the user rating, R played an important role in classifying the users' expertise level. In the case of U_9, although the mapped z-score is the lowest (1 point), the users' expertise can still be promoted through the rating. For U_1 and U_6, the user rating, R will be automatically set to zero since the users did not make any 'answer' post.

Table 4. An example of classification

U_i	q	a	Z	M	R	F	L
U_1	8	0	-2.83	2	0	1	B
U_2	0	5	2.24	4	0	2	I
					1	2.5	I
					2	3	I
					3	3.5	E
					4	4	E
					5	4.5	E
U_3	7	7	0	3	0	1.5	I
					1	2	I
					2	2.5	I
					3	3	I
					4	3.5	E
					5	4	E
U_4	20	10	-1.83	2	0	1	B
					1	1.5	I
					2	2	I
					3	2.5	I
					4	3	I
					5	3.5	E
U_5	5	10	1.29	4	0	2	I
					1	2.5	I
					2	3	I
					3	3.5	E
					4	4	E
					5	4.5	E
U_6	40	0	-6.32	1	0	0.5	B
U_7	0	60	7.75	5	0	2.5	I
					1	3	I
					2	3.5	E
					3	4	E
					4	4.5	E
					5	5	E
U_8	30	30	0	3	0	1.5	I
					1	2	I
					2	2.5	I
					3	3	I
					4	3.5	E
					5	4	E
U_9	110	40	-5.72	1	0	0.5	B
					1	1	B
					2	1.5	I
					3	2	I
					4	2.5	I
					5	3	I
U_{10}	60	150	6.21	5	0	2.5	I
					1	3	I
					2	3.5	E
					3	4	E
					4	4.5	E
					5	5	E

In addition, users are also allowed to flag posts if they find it inappropriate to the topic. After user flags a certain post, the administrator of the portal will be notified to take further action. Through the rating and flagging process, the users of the community are also contributing in giving point to users. As a result, members of the community also contribute to classification of users' level of expertise in the portal.

4. Implementation and results

This paper will use software engineering as its test bed domain. We find that software engineering is an interesting domain as it concerns the creation and maintenance of software application by applying technologies and practices from computer sciences, project management, engineering, application domains, and other fields. The proposed PBaSE method is applied in an existing web portal for software engineers in Malaysia called Malaysian Software Engineering Interest Group (MySEIG). MySEIG was founded in mid 2005 to provide a platform for software engineers to share knowledge, ideas and experience related to software engineering issues [10].

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

The field topics in MySEIG are based on Software Engineering Body of Knowledge or SWEBOK [11] as listed in Table 5. Users are allowed to choose their field of interest from the listed topic. This means each user have a different set of field of interest.

Table 5. Field of interests in MySEIG [11]

No	Topic
1	Software Configuration Management
2	Software Construction
3	Software Design
4	Software Engineering Management
5	Software Engineering Process
6	Software Engineering Tools and Methods
7	Software Maintenance
8	Software Quality
9	Software Requirement
10	Software Testing

The first step of PBaSE method in MySEIG knowledge portal is to calculate the z-score of each user. In order to calculate the z-score, we have to identify the type of posts the created. When users create a new post, they are required to choose the type of post from the dropdown list as in Figure 4.

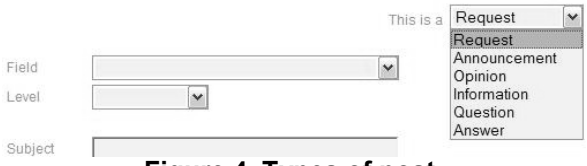

Figure 4. Types of post

There are six types of post the users can choose from. The six types can be categorized into two; 'query' post and 'answer' post. The 'query' post are 'request', 'announcement' and 'question' while 'answer' post are 'opinion', 'information' and 'answer'. The type of post is 'request' by default.

We then can calculate the z-score measures [6]. After the z-score is calculated for every user under a certain field, we will then map the z-score from the scale of 1 to 5 points as shown in Figure 2.

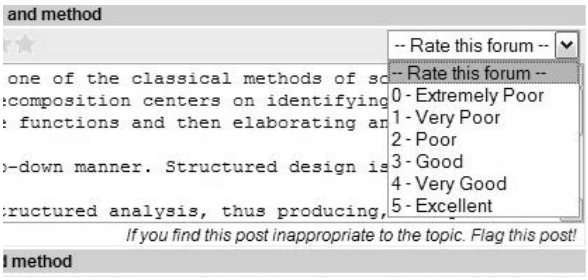

Figure 5. Users' input

PBaSE is a two-way classification method where its users also take part in the classification process. When users view 'answer' post, they are required to rate by the scales: 0 (extremely poor), 1 (very poor), 2 (poor), 3 (good), 4 (very good), 5 (excellent) as shown in Figure 5.

5. Qualitative evaluation

Comparison of PBaSE with existing expertise classification method used in ITTutor.net [4] and Computer Forum [5] is listed in the following aspects:

(i) *Direction of the classification*: Both in ITTutor.net [4] and Computer Forum [5] implement a one-way classification method where the users are not involved in the classification process. On the other hand, PBaSE is a two-way classification method where users are involved in the classification process.

(ii) *Basis of classification*: ITTutor.net [4] and Computer Forum [5] classify its users based on the number of posts they created in the forum while PBaSE classify users based on the average points from the two part of the classification (z-score measure and user rating).

(iii) *Differences in the type of post*: All post treated equally and will be included in the classification process in ITTutor.net [4] and Computer Forum [5]. In contrast, PBaSE uses z-score measures that calculate the distribution of the questions and answers of each user.

(iv) *Competitiveness to be an expert*: There is no competitiveness to be an expert available in ITTutor.net [4] and Computer Forum [5] because the expertise level of the user will not dropped. However, the expertise level of the user can drop if the user stops contributing in the portal. Experts in the portal are always the current top contributors in the portal when using PBaSE method.

978-1-4244-2328-6/08/$25.00 © 2008 IEEE

6. Conclusions and future work

Instead of using the conventional way to classify users based on the number of posts, this paper proposes a two-way classification approach called Point-based Semi-automatic Expertise (PBaSE) classification method. By proposing the PBaSE method, we hope to maximize the capability of SIG knowledge portal for the convenience of its community members.

On the other hand, we have identified that there is a limitation on identifying the type of posts. Based on the current approach, users are required to state the type of post. By integrating the Natural Language Processing (NLP) with PBaSE; users will no longer need to state the type of post since NLP will automatically analyze and identify the type of posts.

7. Acknowledgements

This research is partially supported by research university grant of Universiti Sains Malaysia, 1001/PKOMP/817002.

8. References

[1] Newman, B. (Bo), and Conrad, K. W., "A Framework for Characterizing Knowledge Management Methods, Practices, and Technologies", *The Knowledge Management Forum*, 1999.

[2] Niwa, K., "Towards Successful Implementation of Knowledge-Based Systems: Expert Systems vs. Knowledge Sharing Systems", *IEEE Transactions on Engineering Management*, 37(4), November 1990.

[3] Giarratano, J. and Riley, G., *Expert Systems Principles and Programming: 3rd Edition*, PWS Publishing Co., London, 1998.

[4] ITTutor.net, 2008, http://ittutor.net.

[5] Computer Forum, *Jelsoft Enterprises Ltd.*, 2008, http://www.computerforum.com.

[6] Zhang, J., Ackerman, M. S., and Adamic, L., "Expertise Networks in Online Communities: Structure and Algorithms", *Proceedings of the International World Wide Web Conference WWW 2007*, ACM Press, 2007, pp. 221-230.

[7] Page, L., Brin, S., Motwani, R. and Winograd, T., "The PageRank Citation Ranking: Bringing Order to the Web", *Stanford Digital Library Technologies Project*, 1998.

[8] Kleinberg, J. M., "Hubs, Authorities, and Communities", *ACM Computing Surveys*, 31, U21-U23.

[9] Löser, A. and Tempich, C., "On Ranking Peers in Semantic Overlay Networks", 2005.

[10] MySEIG, 2008, http://www.myseig.org.

[11] Abran, A. *et al.* (eds.), *Guide to Software Engineering Body of Knowledge SWEBOK: 2004 Version*, USA, IEEE Computer Society Press, 2004.