User Reputation Evaluation Using Co-occurrence Feature and Collective Intelligence

Jeong-Won Cha[1], Hyun-woo Lee[1], Yo-Sub Han[2], and Laehyun Kim[3]

[1] Changwon National University, Changwon, Republic of Korea
{jcha, ggamsso}@changwon.ac.kr
[2] Dept. of Computer Science, Yonsei University, Seoul, Republic of Korea
emmous@yonsei.ac.kr
[3] Intelligence and Interaction Research Center, KIST, Seoul, Republic of Korea
laehyunk@kist.re.kr

Abstract. It becomes more difficult to find valuable contents in the Web 2.0 environment since lots of inexperienced users provide many unorganized contents. In the previous researches, people has proved that non-text information such as the number of references, the number of supports, and the length of answers is effective to evaluate answers to a question in a online QnA service site. However, these features can be changed easily by users and cannot reflect social activity of users. In this paper, we propose a new method to evaluate user reputation using co-occurrence features between question and answers, and collective intelligence. If we are able to calculate user reputation, then we can estimate the worth of contents that has small number of reference and small number of support. We compute the user reputation using a modified PageRank algorithm. The experiment results show that our proposed method is effective and useful for identifying such contents.

Keywords: PageRank, User Reputation, Co-occurrence, Collective Intelligence.

1 Introduction

There are many services that share movies, pictures, and knowledge using collective intelligence in the Internet. Examples are Wikipedia, Youtube, Facebook, and GisikiN[1]. These services are based on participation of people's own accord. The most important feature of these services is to find/provide useful data and, thus, each service has its own search engine. However, users often see just-keyword-matched contents that are unrelated and thus unuseful.

Consider 'GisikiN' of NHN that is the largest portal site of Korea. 'GisikiN' is the most popular knowledge sharing service in Korea. Once a user posts a question, other users answer the question similar to Yahoo! Answers. Then, the questioner chooses the best answer and others do thumb-up or thumb-down answers that show the consensus. However, if no best answer is selected by the corresponding questioner, then other users may choose the best answer. In recent years, some studies have attempted to find and explore the quality evaluation of contents [1, 2]. Nevertheless, studies on this problem are still in the early stage of development.

[1] http://kin.naver.com/?frm=nt

A.A. Ozok and P. Zaphiris (Eds.): Online Communities, LNCS 5621, pp. 305–311, 2009.

In this paper, we propose a new method to compute user reputation for finding useful contents. We build a social network based on social activities and the similarity between questions and answers.

The rest of this paper is organized as follows. Section 2 discusses the related work. Our proposed method including co-occurrence feature and collective intelligence is described in Section 3. Various experiments are set up and the results are described in Section 4. We conclude the paper in Section 5.

2 Related Works

There has been a lot researches regarding analysis of hyperlink of documents for computing the importance of documents [3, 4]. Kleinberg found authorized documents using the hyperlink structures of documents related to queries [3]. Brin and Page classify the hyperlinks. They think that if a lot of documents have links point to more important document A and A has a link points to document B, the link of A is more important than others. So they design the PageRank algorithm based on this idea [4]. However, there is a pitfall when a new document is created: even if the new document is very important, it may have a lower rank because it has fewer links than the old ones. Hotho et al. propose FolkRank that is a variant of PageRank [5]. The algorithm is based on the assumption that more authorized authors may write more important tags. They make a network using words, authors, and tags, and, then calculate the importance of documents using the PageRank algorithm.

Note that there are some studies of direct evaluation of contents. For example, [1] evaluates contents using a non-content based method that includes the number of references, the number thumb-ups, and the number of answering comments. [6, 7] use the rate of positive features and negative features for evaluation. [2] evaluates the truth level of documents using content features like keywords, length of document. However, as far as we are aware, there is no known study of user reputation for evaluating user generated contents. We estimate user reputation from social activities and collaborations, and evaluate contents using user reputation.

3 User Reputation Evaluation

3.1 Co-occurrence Feature

We use co-occurrence features to calculate the similarity between question and answer. We use n-grams, which is different from the previous work [8] that uses topic words in blog body and comments to classified spams. Since there are many irregular forms in the web documents, the standard language analytic engines like a part-of-speech tagger do not work well.

First, we collect n-gram from a title or question and answers. Then we calculate the similarity as follows:

$$s(X,Y) = \frac{|X \cap Y|}{|X| + |Y|} \tag{1}$$

where $|X|$ is the number of n-grams of document X. X is a title or question and Y is answers.

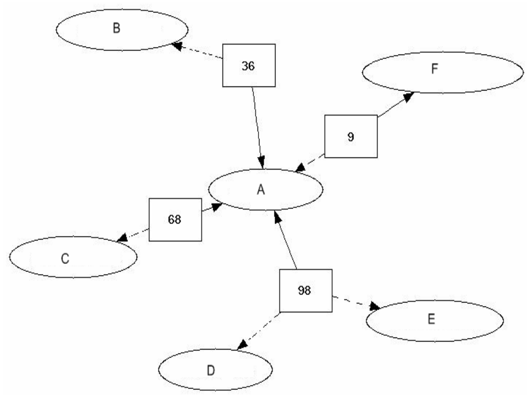

Fig. 1. Social network including questions and users. Rectangles are questions and ovals are users who write answers. Each number in a rectangle and each letter in an oval denote a unique ID for identification.

3.2 Collective Intelligence

We calculate the user reputation using the PageRank algorithm [4] based on collective intelligence. The PageRank algorithm calculates the importance of the document using the number of connections pointed by other documents. We assume that the users who write the questions or answers are nodes like documents in the PageRank algorithm. The proposed algorithm is similar to the PageRank algorithm but is different from the PageRank algorithm in assigning different weights to each link.

There are two different links in our algorithm. One is a 'link from a question to a selected answer (for example, a link from 36 to A in Figure 1)' and the other is a 'link from a question to an unselected answer (for example, a link from 36 to B Figure 1)'. Figure 1 is an example of the social network including questions and users.

The solid lines are 'links that selected to answer' and the dotted lines are 'links that unselected to answer' in Figure 1. Questions have links that point to users who write the answer to the questions. We calculate the user reputation for evaluating the worth of documents as follows:

$$
\begin{aligned}
& UR(p_i) = (1-d) + d \sum_{q_i \in M(p_i)} \frac{f(q_i)}{C(q_i)}, \\
& p_i : user, \\
& q_i : question, \\
& M(p_i) : \text{\# of questions answered by } p_i, \\
& C(q_i) : \text{\# of comments attach to } q_i, \\
& f(q_i) : \begin{cases} \text{if selected answer: } 0.8 \\ \text{otherwise: } 0.2 \end{cases}, \\
& d = 0.85
\end{aligned}
\tag{2}
$$

We observe that a useful answer is not necessarily chosen by a questioner. This leads us to consider unselected answers as well as selected answers. Thus, we set $f(q_i)$ to 0.8 for selected answers and 0.2 for unselected answers through experiments.

4 Experiments

4.1 Test Data

We collect the test data for experiment from 'GisikiN' of NHN. Table 1 shows the information for the data.

Table 1. Data information

# of user	# of questions	# of answers	# of answers/# of questions
20,900	20,588	43,913	2.13

4.2 User Reputation Using Collective Intelligence

We conduct the first experiment using equation (2). Table 2 shows the top-10 ranked users' reputation.

In Table 1, we witness that a user who has high reputation tends to have a more number of selected answers based on equation (2). However, we notice that user6 (U6) ranks at 6th although he has less selected answers and more unselected answers in comparison to the others. This is because we assign weight 0.2 to unselected answers uniformly.

To improve this weakness, we compute content similarity using n-gram co-occurrence features at question and answer. We use the title and the body of questions and answers for the similarity. For selecting an appropriate n value, we calculate the similarity for different n. Table 3 shows that the selected answers have higher

Table 2. We conduct experiment 1 using equation (2). Note that 'user' is a person who writes an answer, 'reputation' is user reputation, 'selection' is the number of selected answers, and 'non-selection' is the number of unselected answers by users.

Ranking	User ID	Reputation	Selection	Non-selection
1	U1	0.0075	771	194
2	U2	0.0048	350	518
3	U3	0.0044	361	57
4	U4	0.0043	397	107
5	U5	0.0039	378	234
6	*U6	0.0039	258	828
7	U7	0.0037	382	278
8	U8	0.0035	332	384
9	U9	0.0031	382	125
10	U10	0.0029	254	259

Table 3. The question-answer similarity according to n-gram size. From the result of experiment, we use 2-gram to compute similarity between question and answer.

n-gram size	2	3	4
Similarity between selected answer and question	**0.6764**	0.4081	0.1818
Similarity between non-selected answer and question	0.4089	0.2634	0.1519

similarity value than the unselected answers' using 2-gram size. This leads us to use bi-gram for calculating the similarity between them.

After calculating the similarity, we modify the user reputation equation as follows:

$$UR(p_i) = (1-d) + d \sum_{q_i \in M(p_i)} \frac{f(q_i) \times s(q_i)}{C(q_i)}, \tag{3}$$

where, $s(q_i)$ is the question-answer similarity using co-occurrence features. We conduct additional experiment using equation (3). At experiment 2 in Table 4, we note that U1's reputation is increased because of using the question-answer similarity whereas U6's reputation is decreased compared with experiment 1. This shows question-answer similarity is effective compared with the non-text features like the number of answers.

We introduce another good feature for evaluating contents from a social network: the number of recommendations by anonymous readers. Regardless of questioner choice of answers, a good answer receives many recommendations. We apply this observation in equation (4): we use the ratio of recommendation instead of the number of recommendations for normalization. We use different weights according to selected answers, unselected answers, and self-answer. Because the self-answer can

Table 4. We conduct experiment 2 using equation (3). Note that 'user' is a person who writes an answer, 'reputation' is user reputation, 'selection' is the number of selected answers, and 'non-selection' is the number of unselected answers by users.

Ranking	User ID	Reputation	Selection	Non-selection
1	U1	0.0114	771	194
2	U2	0.0078	350	518
3	U4	0.0075	397	107
4	U5	0.0057	378	234
5	U7	0.0056	382	278
6	U3	0.0054	361	57
7	U9	0.0053	382	125
8	U8	0.0051	332	384
9	*U6	0.0049	258	828
10	U99	0.0049	237	138

fabricate user's reputation, we assign a low rate to it. From the results, we know that the rank of a user who has many unselected answers and low recommendation, for instance U6, falls far behind compared with the experiment 2 case.

$$UR(p_i) = (1-d) + d \sum_{q_i \in M(p_i)} \frac{f(q_i) \times s(q_i) \times r(q_i)}{C(q_i)},$$
$$r(q_i): \begin{cases} \text{selected answer : recommeded ratio} \times 0.6 \\ \text{unselected answer : recommended ratio} \times 0.3 \\ \text{self-answer : recommended ratio} \times 0.1 \end{cases} \tag{4}$$

Table 5 shows the result of experiment 3. Note that U6 does not exist at the top-10 ranked users. Newly, U12 comes into the lists.

Table 5. We conduct experiment 3 using equation (4). Note that 'user' is a person who writes an answer, 'reputation' is user reputation, 'selection' is the number of selected answers, and 'non-selection' is the number of unselected answers by users.

Rank	User ID	Reputation	Selection	Non-selection
1	U1	0.0114	771	194
2	U4	0.0075	397	107
3	U2	0.0073	350	518
4	U5	0.0056	378	234
5	U7	0.0055	382	278
6	U3	0.0054	361	57
7	U9	0.0053	382	125
8	U8	0.0049	332	384
9	U99	0.0049	237	138
10	U12	0.0043	264	67

5 Conclusions

Web 2.0 emphasizes user participation. The participation of user in a social network is effective criterion of user reputation. In this paper, we propose a new method to calculate user reputation using co-occurrence features and collective intelligence for selecting good answer given the questions. We consider the 'GisikiN' of NHN as a sample site. We define a social network using questions and users who write answers in GisikiN. We conduct experiments on test data from GisikiN, and the results show the effectiveness of our proposed method. The good performance of the proposed method is useful to evaluate the answer generated by users given questions. Developing extended method adding non-text features will be our future work.

Acknowledgements. This work was supported by the IT R&D program of MKE/IITA 2008-S-024-01.

References

1. Jeon, J., Croft, W.B., Lee, J.H., Park, S.: A Framework to Predict the Quality of Answers with Non-Textual Features. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, USA, pp. 228–235 (2006)
2. Lee, J., Song, Y., Rim, H.: Quality Prediction of Knowledge Search Documents Using Text-Confidence Features. In: Proceedings of Hangul and Cognitive Language Technology 2008, pp. 62–67 (2007) (in Korean)
3. Kleinberg, J.: Authoritative Sources in a Hyperlinked Environment. Journal of the ACM 46(5), 604–632 (1999)
4. Brin, S., Page, L.: The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 30, 107–117 (1998)
5. Hotho, A., Jaschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. The Semantic Web: Research and Applications 4011, 411–442 (2006)
6. Zhu, X., Gauch, S.: Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web. In: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 288–295 (2000)
7. Zhou, Y., Croft, W.B.: Document quality models for web ad hoc retrieval. In: Proceedings of the 14th ACM international conference on Information and knowledge management, pp. 331–332 (2005)
8. Jeon, H.-W., Rim, H.-C.: A Comment Spam Filter System based on Inverse Chi-Square Using of Co-occurrence Feature Between Comment and Blog Post. In: Proceedings of HCLT 2007, pp. 122–127 (2007) (in Korean)