

COMPUTING USER REPUTATION IN A SOCIAL NETWORK OF WEB 2.0

Yo-Sub HAN

*Department of Computer Science, Yonsei University
Seoul, Republic of Korea
e-mail: emmous@cs.yonsei.ac.kr*

Laehyun KIM

*Intelligence and Interaction Research Center, KIST
Seoul, Republic of Korea
e-mail: laehyunk@kist.re.kr*

Jeong-Won CHA

*Department of Computer Engineering, Changwon National University
Changwon, Republic of Korea
e-mail: jcha@changwon.ac.kr*

Communicated by Michal Laclavík

Abstract. In the Web 2.0 era, people not only read web contents but create, upload, view, share and evaluate all contents on the web. This leads us to introduce a new type of social network based on user activity and content metadata. We notice that we can determine the quality of related contents using this new social network. Based on this observation, we introduce a user evaluation algorithm for user-generated video sharing website. First, we make a social network of users from video contents and related social activities such as subscription, uploading or favorite. We then use a modified PageRank algorithm to compute user reputation from the social network. We re-calculate the content scores using user reputations and compare the results with a standard BM25 result. We apply the proposed approach to YouTube and demonstrate that the user reputation is closely related

to the number of subscriptions and the number of uploaded contents. Furthermore, we show that the new ranking results relied on the user reputation is better than the standard BM25 approach by experiments.

Keywords: User reputation, social network, Web 2.0, YouTube

1 INTRODUCTION

In the early 1990s web, which is often called Web 1.0, most people just read and watch online contents such as web documents and videos that are provided by a small number of special people – webmasters. The information flow in the web is similar to the traditional publishing process: A small number of publishers provide contents to a large number of readers. However, since the mid-1990s, the web has changed drastically: Web 2.0 has appeared [15]. Web 2.0 does not refer to any technical specification of the World Wide Web but rather to changes of ways how people use the Web. The slogan for Web 2.0 is participation, sharing and openness. End-users of the Web not only read contents but upload and share own contents. Once content is uploaded on the Web, other end-users give feedback by rating or commenting, and modify original contents creatively, which give rise to a new content. Therefore, there is no longer clear distinction between web content providers and consumers, and the information flow is now bidirectional. Blogs, Wikipedia, YouTube¹ and Facebook² are an example of Web 2.0 platform. The bidirectional interactions among users naturally form a user network in a closed platform. We can think of a closed platform as a community and a user network as a social network of the community.

Consider YouTube for example. YouTube is one of the most popular video sharing web communities. Once a user uploads a video that s/he wants to share, other users can view, subscribe, add to favorite, comment and rate the video. A user interacts with other users via a video indirectly. Furthermore, these interactions give additional information that helps estimate the values of the corresponding video content. For example, a video may have several comments, ratings, favorites and subscriptions by other users. We call these interactions social activity of users. Note that YouTube is a semi-closed platform: any users can watch videos but users have to become a member of YouTube to perform other social activities such as to subscribe, rate or comment.

Next, we explain how the social activity helps estimate user reputation using the following scenario in YouTube. Figure 1 depicts an example of social activities in YouTube. A user A makes her own UGC (user generated content) and uploads it into YouTube. Then, other users execute various comments available in YouTube after watching it. For instance, B leaves a comment, C gives a star score, D adds

¹ <http://www.youtube.com>

² <http://www.facebook.com>

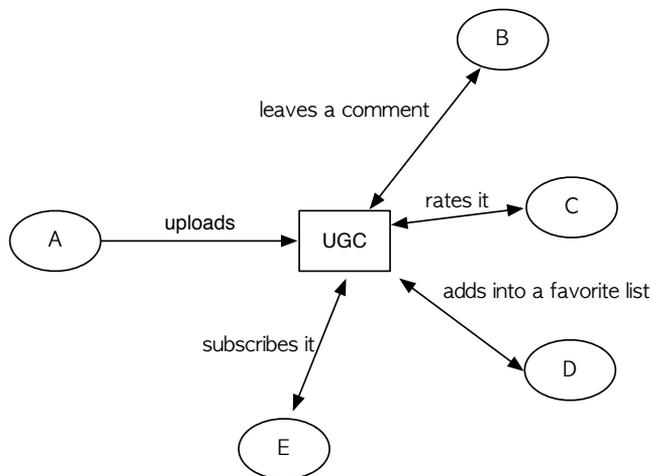


Fig. 1. An example of social activities in YouTube. A, B, C, D and E are users and A is a video uploader.

the contents into her favorite list and E subscribes all of A's contents. Note that these social activities are bidirectional. Namely, users watch a content and response it with various activities. From the social activities via a video, we can create a network among users and it becomes a social network in YouTube. In other words, A and B become a neighborhood of each other. Moreover, since D adds the content into her favorite list, it implies that the content is worth to replay sometime later. Therefore, the quality of the content must be high and, thus, the uploader A is also very reliable for making quality UGCs. Based on this observation, we can estimate user reputation in the new social network. We compute user reputation of a social network from their social activities. Notice that our social network is different from social networking services such as Facebook or Cyworld³. The current social networking services ask users to explicitly set up their social network using a friends list whereas we build a social network of users implicitly based on their social activities with respect to contents.

Once we establish a social network of users, then we compute reputation of each user in the community. We notice that writing a comment is neutral compared with adding a video in his/her favorite list. Namely, one may leave a negative comment or a meaningless comment. On the other hand, adding a video in a favorites list clearly indicates that the video is valuable. Therefore, we consider only important social activities.

We describe related work in Section 2 and introduce our user reputation algorithm in Section 3. Then, in Section 4, we show experiment results and analyze

³ <http://www.cyworld.com>

user reputation parameters based on the results. We show future direction of our method and conclude the paper in Section 5.

2 BACKGROUNDS

2.1 Ranking Algorithms

Given a set of elements, a ranking algorithm computes a relative importance of all elements in the set and orders elements according to the importance. This makes users to find important elements at the top. Web page ranking algorithms are based on the content analysis and the link analysis of web pages. Examples are PageRank [6], TrustRank [10], Anti-Trust Rank [14] and XRank [20]. The web page link structure and the social network in a web community are similar except for the fact that there are more types of links in social network compared to web pages. The PageRank algorithm calculates the importance of a page as the contribution from connecting nodes with out-links in the page. Notice that the algorithm does not analyze the content of page itself and solely relies on the link information among web pages. TrustRank filters out spams from the searching process by selecting some trustful seed sites and processing the link structure, which is the same to the PageRank approach, from the seed sites. Anti-Trust Rank propagates TrustRank in a reverse direction: it starts from a set of seed spam sites instead of good sites. The Anti-Trust Rank algorithm is based on the observation that spam sites share many garbage keywords and links among them to have a high page rank score. Recently, Radovanović et al. [16] proposed a web search engine using text categorization that enhances the order of search results. While some algorithms use link analysis to evaluate the importance of pages, XRank takes a different approach: it considers the site popularity and importance before calculating the importance of pages. For personalized web search engines, Bieliková and her co-authors [4, 5] considered ontology-based and rule-based methods for user model. Note that these ranking algorithms work well for web pages since web pages often have several in-/out-links. However, user-generated video contents may not have explicit link connection between contents. Because of this structural difference, the known link analysis algorithms are not directly applicable for UGCs. Moreover, there are several new types of data in UGCs that cannot be found in web pages. For example, various interactions between content uploaders and viewers can help evaluate the corresponding content.

2.2 User Reputation

There is considerable research on the e-commerce reputation, especially the effect of online reputation about a trader's trust [2] and auction price [19]. Dellarocas [8] introduced a binary reputation model for analyzing market places. In this model, he uses reputation for determining whether or not sellers advertise truthfully. Buyers may exercise some leniency when rating sellers, which needs to be compensated

by corresponding strictness when judging sellers' feedback profiles. Resnick and Zeckhauser [17] described in detail the reputation system in eBay. They compute users reputation after collecting user ratings. They rely on several principles such as the simple summation model in eBay⁴ and the Bayesian model [13] to compute user reputation from user ratings. In order to avoid the unfair factor caused by subjective ratings, Dellarocas [7] proposed an immunization method against unfair ratings and discriminatory behavior. Note that these methods are based on a cross rating of participants; namely, they compute reputation scores based on ratings using aggregative approaches. Meanwhile, Sabater and Sierra [18] have applied social network analysis in reputation system of the multi-agent system.

2.3 Collective Intelligence and Reliability Analysis

Collective intelligence is an intelligence formed from the collaborations of many individuals. Collective intelligence often appears in decision making. Given a data, we consider various user feedbacks such as reading, leaving a comment or showing one's like or dislike. By collecting enough feedbacks, we evaluate the data. Because of mass feedbacks from users, we can estimate the true value. Thus, the principle of collective intelligence is the ability to harness data created by people in a variety of ways. Google is a good example: they rely on the fact that people make hyperlinks only for important keywords and pages. Thus, the algorithm uses the link information and builds up the rank among pages. Gliner et al. [9] showed how measurement reliability and measurement validity are used to determine the research validity. They defined reliability to be the consistency of a particular test instrument. Note that the correlation coefficient is often used as a measure of consistency. Bennet et al. [3] described reliability as an association of credibility, trustworthiness and dependability. Thus, reliability is a quantified measure of uncertainty of events. The reliability issue becomes crucial in online communities because of anonymity of users. This makes the task to identify reliable users very difficult. Researchers use a statistical approach for identifying reliable users. For example, Agarwal et al. [1] analyzed blog activities such as posting and commenting to recognize reliable authors of blog articles. However, a simple statistical approach has a pitfall. For instance, a user may post many useless articles and leave meaningless comments to improve his/her reliability. Therefore, we also need to consider the reliability of contents before computing related user reliability. We make the reliability of a user as a user reputation in an online community.

2.4 YouTube

The link connection in UGCs is different from the link connection in web pages. The link in a web page is a hyperlink defined by <a> tag between two pages. On the other hand, the link in UGCs can be of different types such as rating, commenting

⁴ <http://www.ebay.com>

or reviewing. Thus, we can obtain different information from different link types in UGCs. Another difference between UGCs and web pages is that UGCs are often related to users who can be creators, uploaders or reviewers but web pages are not. We take YouTube as an example system in this paper. YouTube is a famous UGC sharing site. In the system, UGCs are mainly videos. We separate users and video contents:

1. Video content: For each video content in YouTube, the system has several information as illustrated in Figure 2. For example, rating from 1 star to 5 stars, comments, favorites, content sharing to other social network website such as MySpace⁵, Facebook, del.icio.us⁶, and Digg⁷, and honors/awards (most viewed or top rated).
2. User: There are two types of users. One is a content uploader and the other is a reviewer. However, these two groups are not mutually exclusive. Namely, one can upload a video content and watch another video. We define an uploader to be a user who uploads a video content and a reviewer to be a user who watches a video and may give feedback such as commenting or ratings.
 - (a) Uploader: an uploader has a channel or a personal page that can be accessed by other users and, thus, builds a connection with other users by adding them as friends. Other users can also subscribe to one or several channels and the subscription creates a social connection.
 - (b) Reviewer: a reviewer contributes to the measurable-scoring scheme by giving comments to channel, comments to contents, favorites, ratings and scoring content comments.

Note that there are many other social activities that can be used for connecting users in other UGC sharing sites. Here we only consider available connection in YouTube. In the next section, we present an algorithm that computes user reputation in a social network from YouTube based on social activities.

3 PROPOSED METHOD

3.1 Building Social Network

We use YouTube, which is the most famous video sharing site, as our test platform. From YouTube, we obtain users and build up a social network based on their social activities. We do not distinguish users and contents while building social network and computing reputation: We regard both users and contents as just nodes. The reason is that the more a user gets high reputation, the more we trust his/her contents, and vice versa. However, we make difference edges according to relations

⁵ <http://www.myspace.com>

⁶ <http://del.icio.us>

⁷ <http://digg.com>



Fig. 2. An example of a YouTube video content and related information

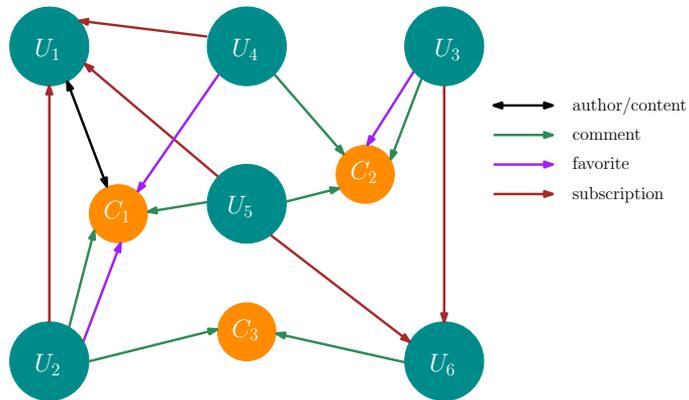


Fig. 3. An example of a social network based on video contents and related social activities from Figure 1. Note that C_i for $i = 1, 2, 3$ denote video contents and U_i for $i = 1, 2, 3, 4, 5, 6$ denote users.

that are author/contents, comments, favorite and subscription. When someone builds and uploads any contents, there are bi-directional author/content relations. When someone gives a comment to a content, there is a comment relation, which, in this case, is a directional relation from a user to a content. If a user puts a content into his favorites list, there is a favorite relation. If a user subscribes another user, then it creates a subscription relation. Figure 3 gives an example of such social network in YouTube.

Then, we compute user reputation of all users in the network. We then improve the previous result⁸ and investigate the correlations between user reputations and social activities in the new social network. We, moreover, re-calculate content rating scores using the obtained user reputations. Finally, we demonstrate the usefulness of the new ranking method combined with user reputation by experiments and user study.

3.2 Computing User Reputation

One of many reasons for the success of the PageRank algorithm [6] is that the algorithm can determine the importance of a web page that is a part of huge web. The algorithm is based on the assumption that a page is important if it is linked by other important pages. We find a similar phenomenon in YouTube. Every day, there are more than 200 000 new video contents uploaded and many new users joined. Furthermore, there are numerous interactions between users via contents. We observe that if a video content is interesting and thus valuable, then it is often added into a favorites list of another user and may have many comments. Moreover, if a user has uploaded many valuable videos, then other users tend to subscribe the user's channel and thus give more frequent access to the videos in the channel. In other words, a user who has many subscriptions (or links) is popular and may have many valuable contents. This leads us to apply the PageRank algorithm for the social network of YouTube that we have established. First, we obtain a social network based on social activities as illustrated in Figure 3 and assign value 1 to each edge of the network for initialization. Then, we run a modified PageRank algorithm using the following formula:

$$UR(U_i) = d + (1 - d) \times \sum_{j=M(U_i)} \frac{w(U_j) \times UR(U_j)}{C(U_j)}, \quad (1)$$

where

- $w(U_j) = \begin{cases} 0.35 & \text{if link is the subscription link} \\ 0.3 & \text{if link is the uploaded content link} \\ 0.2 & \text{if link is the favorite link} \\ 0.15 & \text{otherwise} \end{cases}$
- U_i denotes a node in the network,

⁸ A preliminary approach of computing user reputation was presented in HCI (12) [11].

- $d = 0.15$ is a damping factor,
- $M(U_i)$ is the number of links from U_i and
- C is the number of outgoing links.

Once the algorithm is completed, we have a score for each node in the network. We define this score to be a reputation (or reliability) in a community. As shown in our formula, we assign different weights to different types of edges by experiment. This is because one link might be more valuable than another link. For instance, subscription link is more valuable than, say, comment link since a user only subscribes other users only when s/he thinks it is worth whereas s/he gives a (negative) comment to any videos.

Since we use both users and videos as nodes in the network as depicted in Figure 3, we also have a score for videos. We regard this score as a popularity score. However, we do not consider content popularity here. We will address this issue in a different paper. Therefore, we have computed scores for all nodes and we are only interested user nodes and the corresponding scores.

4 EXPERIMENT RESULTS

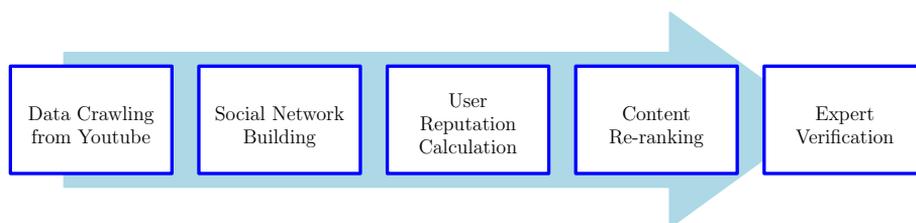


Fig. 4. The experiment process steps

Figure 4 depicts the experimental process. First, we have collected about 600 000 videos and 625 000 users from YouTube and created a social network based on social activities and related content links. Then, we have run the PageRank algorithm using the formula given in Section 3 and have computed score for users and videos. Table 1 shows the top 20 users with high user reputation.

Table 4 shows the top 10 000 high reputation users and their number of subscriptions. Note that highly reputed users have much more subscriptions than the other users. Moreover, the total sum of subscriptions of reputed users is much bigger than the sum of all other users. This implies that we can identify a relatively small number of users with high reputation among numerous users in community.

We also notice that user reputation is closely related with the number of uploaded contents. Thus, we compare correlation between each pair of features in Figure 5 and established the following result in Table 2.

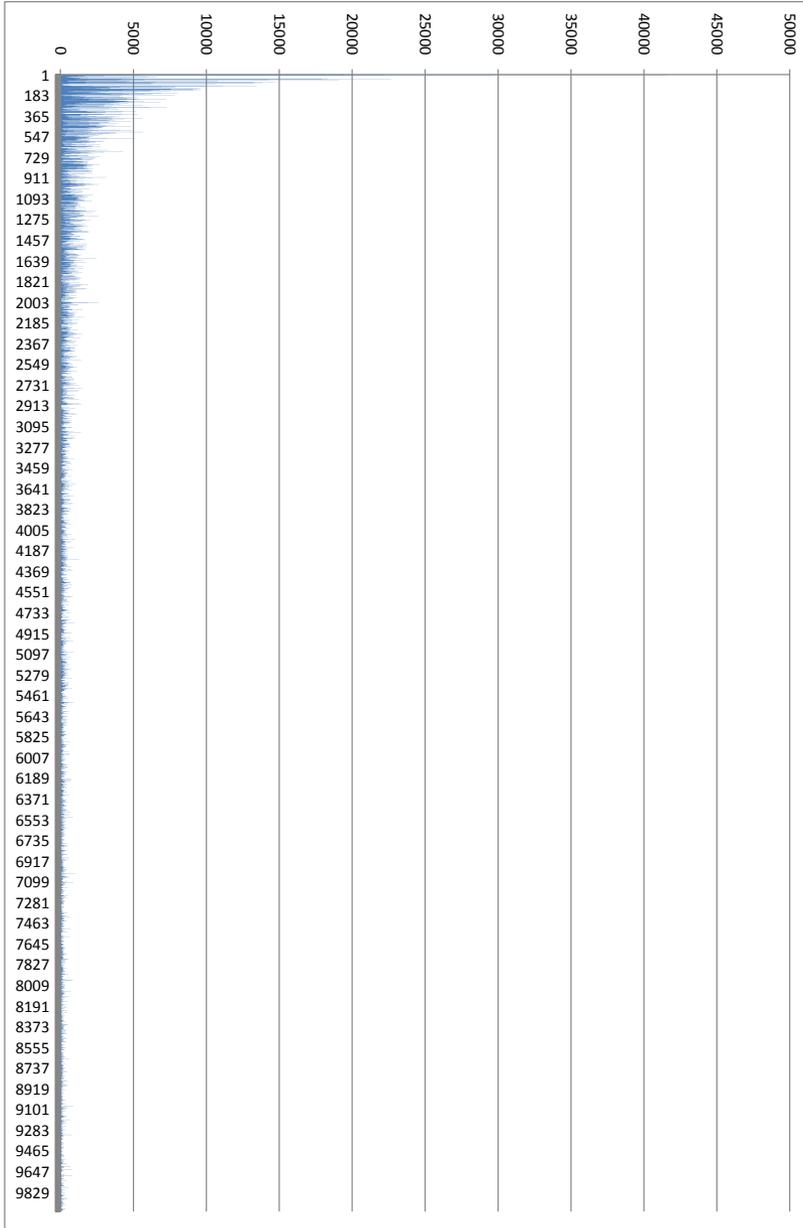


Fig. 5. The x -axis is the ranking according to user reputation and the y -axis is the number of subscriptions of each user

| ID | # of subscriptions | AVG rating | # of contents | UR |
|---------------------|--------------------|------------|---------------|------|
| nigahiga | 45 301 | 4.80 | 52 | 4.31 |
| universalmusicg | 19 452 | 4.78 | 931 | 4.29 |
| machinima | 22 145 | 4.56 | 850 | 3.03 |
| smosh | 41 612 | 4.65 | 54 | 2.31 |
| sonybmj | 3 010 | 4.84 | 915 | 1.99 |
| NBA | 4 445 | 4.66 | 878 | 1.92 |
| BarackObamadc | 5 442 | 4.65 | 814 | 1.82 |
| CBS | 2 853 | 4.52 | 854 | 1.74 |
| lockergnome | 5 493 | 4.49 | 887 | 1.73 |
| NHLVideo | 1 568 | 4.54 | 880 | 1.68 |
| Fred | 29 619 | 4.61 | 22 | 1.64 |
| CSPANJUNKIEdc | 1 217 | 4.65 | 877 | 1.62 |
| Journeymanpicture | 1 217 | 4.55 | 890 | 1.59 |
| Google | 3 010 | 4.07 | 791 | 1.58 |
| Rrden | 51 | 4.46 | 891 | 1.56 |
| AlJazeeraEnglish | 1 704 | 4.64 | 830 | 1.56 |
| travelandtransition | 21 | 4.37 | 888 | 1.53 |
| malaysiakini | 97 | 3.97 | 875 | 1.52 |
| NationalGeographic | 5 954 | 4.62 | 671 | 1.48 |
| CharlieRose | 588 | 4.78 | 849 | 1.48 |

Table 1. A list of top 20 users with high user reputation scores. AVG rating is the average rating of all contents made by the corresponding user and UR denotes user reputation.

| UR vs upload | UR vs subscription | UR vs rating | subs vs upload |
|--------------|--------------------|--------------|----------------|
| 0.84 | 0.61 | -0.01 | 0.17 |

Table 2. Correlation between user reputation and other factors in Table 1

Note that although the subscription link has a higher weight (0.35) than the upload link (0.3) in the formula given in Section 3, the correlation between user reputation and the number of uploaded contents is higher. On the other hand, the average rating is not quite related with user reputation. This shows that the current rating system does not consider user reputation since it gives the same importance to every user no matter whether user reliability is good or bad.

Based on the user reputation that we have computed, we re-calculate a rating score for a content based on each rater's user reputation and a given score. We do not have any standard test set for video contents. We use the BM25 search model [12] as a comparison model: We have implemented a new search engine using the BM25 model. The engine ranks contents based on titles, tags and descriptions of video contents. We perform 10 queries to two systems and 9 experts give 10 grade points to the top 5 contents from each system. For example, we run 'ipod' to two systems (PM and BM). Then, two systems give results using their own ranking method. The expert give 10 grade points to the top 5 contents from each system. That is

10 is invaluable and 0 is useless. Then we calculate the score using the following equation:

$$\text{score} = \frac{1}{Q} \sum_{i=1}^Q \sum_{j=1}^5 \left(\frac{s_j}{r_j} \right)^2, \quad (2)$$

where Q is the number of queries (in this paper we use 10 queries.), s_j is the point given by experts to each content, and r_j is the rank of the content. This equation is designed according to the concept that high points and high rank are more important. Table 3 shows each point according to the queries.

Table 4 shows final scores given by experts to each model. For example, from the scores by expert ‘A’ in Table 3, we compute the score for the proposed method using Equation (2) as follows:

$$\begin{aligned} \text{score} &= \frac{1}{10} \left\{ \begin{aligned} &\left\{ \left(\frac{8}{1} \right)^2 + \left(\frac{1}{2} \right)^2 + \left(\frac{3}{3} \right)^2 + \left(\frac{5}{4} \right)^2 + \left(\frac{5}{5} \right)^2 \right\} && \text{(query: anima.)} \\ &+ \left\{ \left(\frac{2}{1} \right)^2 + \left(\frac{1}{2} \right)^2 + \left(\frac{5}{3} \right)^2 + \left(\frac{5}{4} \right)^2 + \left(\frac{5}{5} \right)^2 \right\} && \text{(query: naruto)} \\ &\dots \\ &+ \left\{ \left(\frac{0}{1} \right)^2 + \left(\frac{2}{2} \right)^2 + \left(\frac{2}{3} \right)^2 + \left(\frac{2}{4} \right)^2 + \left(\frac{2}{5} \right)^2 \right\} && \text{(query: iphone)} \\ &+ \left\{ \left(\frac{0}{1} \right)^2 + \left(\frac{3}{2} \right)^2 + \left(\frac{6}{3} \right)^2 + \left(\frac{0}{4} \right)^2 + \left(\frac{6}{5} \right)^2 \right\} && \text{(query: obama)} \\ &\} \\ &= 30.7 \end{aligned} \right. \end{aligned}$$

We can find from Table 4 that PM has better performance than BM: The high score is better than the low score. Notice that experts ‘A’, ‘C’, ‘F’, ‘G’ and ‘H’ have ruled in favor of PM.

We observe that the difference between the proposed (PM) and the BM25 approach in Table 4 is small. This is because the social network that we use is not big enough. For instance, we have collected about 600 000 videos contents and related information from YouTube and this is very small compared with the whole contents available on YouTube. We expect that the large amount of contents and users can improve the searching performance. Note that the BM25 approach relies on the text information such as tags, body and title descriptions and YouTube has enough of such text information for its UGCs. On the other hand, there are many UGC sites that do that have such information available. For those UGC sites, our approach can improve the searching quality by using social network and user reputation.

5 CONCLUSIONS

People produce and upload a lot of user-generated contents in Web 2.0 and it is beyond the computing power to process each content for evaluating the value of each one correctly. On the other hand, we can use the help from humans and utilize

| | anima. | | naruto | | myspace | | guitar | | lyrics | | ipod | | apple | | google | | iphone | | obama | |
|---|--------|----|--------|----|---------|----|--------|----|--------|----|------|----|-------|----|--------|----|--------|----|-------|----|
| | PM | BM | PM | BM | PM | BM | PM | BM | PM | BM | PM | BM | PM | BM | PM | BM | PM | BM | PM | BM |
| A | 8 | 1 | 2 | 3 | 0 | 0 | 8 | 7 | 8 | 5 | 4 | 2 | 1 | 2 | 6 | 6 | 0 | 2 | 0 | 3 |
| | 1 | 5 | 1 | 3 | 0 | 0 | 0 | 3 | 8 | 4 | 0 | 2 | 2 | 8 | 2 | 2 | 2 | 5 | 3 | 3 |
| | 3 | 8 | 5 | 6 | 0 | 0 | 6 | 7 | 6 | 1 | 6 | 3 | 2 | 3 | 1 | 2 | 2 | 2 | 6 | 5 |
| | 5 | 3 | 5 | 3 | 0 | 0 | 6 | 7 | 4 | 1 | 0 | 5 | 5 | 2 | 0 | 2 | 2 | 2 | 0 | 5 |
| | 5 | 1 | 5 | 3 | 0 | 5 | 8 | 3 | 1 | 4 | 2 | 1 | 5 | 2 | 0 | 2 | 2 | 0 | 6 | 3 |
| B | 8 | 0 | 0 | 6 | 0 | 5 | 10 | 10 | 10 | 7 | 0 | 10 | 0 | 5 | 5 | 8 | 0 | 8 | 0 | 5 |
| | 3 | 10 | 0 | 6 | 0 | 5 | 0 | 1 | 10 | 10 | 8 | 8 | 3 | 10 | 5 | 10 | 9 | 10 | 5 | 10 |
| | 10 | 5 | 3 | 5 | 0 | 5 | 10 | 10 | 0 | 0 | 10 | 9 | 10 | 7 | 0 | 5 | 10 | 8 | 0 | 10 |
| | 5 | 1 | 3 | 0 | 0 | 0 | 10 | 5 | 5 | 0 | 0 | 5 | 10 | 7 | 0 | 5 | 0 | 5 | 0 | 0 |
| | 5 | 0 | 0 | 6 | 0 | 10 | 10 | 1 | 0 | 10 | 9 | 1 | 10 | 8 | 0 | 3 | 8 | 0 | 10 | 10 |
| C | 7 | 8 | 1 | 4 | 1 | 1 | 9 | 5 | 6 | 3 | 4 | 3 | 4 | 8 | 5 | 3 | 2 | 4 | 1 | 6 |
| | 1 | 5 | 6 | 4 | 3 | 1 | 6 | 5 | 7 | 2 | 3 | 1 | 3 | 4 | 5 | 2 | 3 | 2 | 2 | 5 |
| | 2 | 9 | 7 | 5 | 1 | 1 | 9 | 5 | 3 | 1 | 5 | 2 | 5 | 1 | 3 | 2 | 5 | 2 | 7 | 1 |
| | 8 | 4 | 7 | 3 | 4 | 5 | 3 | 6 | 3 | 3 | 2 | 4 | 3 | 1 | 6 | 2 | 2 | 2 | 3 | 1 |
| | 8 | 3 | 6 | 2 | 9 | 3 | 9 | 2 | 2 | 2 | 3 | 2 | 7 | 2 | 3 | 1 | 2 | 1 | 1 | 2 |
| D | 9 | 0 | 5 | 8 | 0 | 0 | 9 | 8 | 8 | 6 | 3 | 10 | 2 | 8 | 10 | 6 | 1 | 7 | 0 | 5 |
| | 7 | 9 | 9 | 8 | 0 | 0 | 0 | 6 | 8 | 5 | 4 | 6 | 4 | 9 | 7 | 8 | 7 | 6 | 1 | 4 |
| | 5 | 10 | 5 | 10 | 0 | 0 | 9 | 10 | 7 | 0 | 9 | 8 | 6 | 8 | 2 | 7 | 6 | 2 | 5 | 10 |
| | 7 | 9 | 6 | 8 | 0 | 0 | 4 | 8 | 5 | 3 | 6 | 7 | 7 | 8 | 0 | 7 | 9 | 4 | 0 | 9 |
| | 8 | 0 | 7 | 8 | 0 | 0 | 10 | 6 | 5 | 5 | 10 | 9 | 8 | 5 | 2 | 8 | 8 | 0 | 10 | 4 |
| E | 3 | 7 | 1 | 3 | 3 | 2 | 8 | 8 | 4 | 6 | 1 | 4 | 1 | 7 | 7 | 4 | 4 | 5 | 1 | 2 |
| | 4 | 6 | 1 | 3 | 4 | 3 | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 5 | 4 | 4 | 4 | 4 | 4 | 3 |
| | 2 | 7 | 4 | 7 | 6 | 3 | 5 | 8 | 7 | 4 | 4 | 3 | 4 | 3 | 4 | 5 | 7 | 3 | 2 | 4 |
| | 2 | 5 | 4 | 4 | 4 | 5 | 3 | 5 | 2 | 3 | 3 | 3 | 6 | 3 | 3 | 3 | 4 | 6 | 6 | 4 |
| | 3 | 6 | 4 | 4 | 4 | 4 | 3 | 1 | 3 | 4 | 3 | 7 | 4 | 4 | 3 | 4 | 3 | 3 | 2 | |
| F | 3 | 0 | 10 | 7 | 0 | 0 | 10 | 7 | 5 | 9 | 5 | 3 | 5 | 5 | 6 | 2 | 2 | 3 | 0 | 6 |
| | 4 | 4 | 5 | 7 | 0 | 0 | 0 | 5 | 5 | 3 | 5 | 2 | 5 | 3 | 3 | 3 | 3 | 5 | 3 | |
| | 2 | 4 | 5 | 8 | 0 | 0 | 8 | 7 | 9 | 8 | 0 | 4 | 3 | 3 | 0 | 4 | 7 | 4 | 3 | 2 |
| | 4 | 5 | 6 | 5 | 0 | 0 | 5 | 4 | 3 | 3 | 3 | 4 | 2 | 2 | 0 | 2 | 6 | 3 | 0 | 5 |
| | 8 | 0 | 5 | 7 | 0 | 5 | 10 | 5 | 0 | 3 | 0 | 3 | 7 | 1 | 4 | 2 | 3 | 0 | 3 | 5 |
| G | 9 | 2 | 6 | 7 | 8 | 5 | 8 | 9 | 10 | 7 | 5 | 9 | 5 | 7 | 10 | 6 | 1 | 4 | 1 | 6 |
| | 4 | 5 | 5 | 7 | 8 | 5 | 2 | 8 | 10 | 8 | 1 | 6 | 1 | 10 | 6 | 8 | 5 | 7 | 3 | 6 |
| | 4 | 6 | 8 | 10 | 9 | 5 | 10 | 9 | 10 | 3 | 10 | 5 | 10 | 8 | 0 | 6 | 8 | 6 | 8 | 8 |
| | 6 | 8 | 8 | 7 | 10 | 6 | 8 | 6 | 10 | 5 | 2 | 9 | 10 | 5 | 0 | 6 | 2 | 5 | 1 | 8 |
| | 9 | 2 | 6 | 10 | 10 | 8 | 10 | 4 | 2 | 7 | 5 | 5 | 10 | 5 | 0 | 6 | 8 | 2 | 9 | 7 |
| H | 3 | 8 | 5 | 6 | 7 | 1 | 8 | 8 | 5 | 5 | 3 | 5 | 3 | 3 | 5 | 1 | 1 | 1 | 7 | 6 |
| | 3 | 7 | 5 | 5 | 5 | 1 | 7 | 7 | 5 | 5 | 4 | 5 | 4 | 4 | 1 | 1 | 4 | 4 | 7 | 7 |
| | 1 | 7 | 6 | 6 | 4 | 1 | 8 | 8 | 6 | 6 | 5 | 1 | 5 | 3 | 3 | 3 | 5 | 5 | 6 | 5 |
| | 1 | 6 | 5 | 4 | 6 | 5 | 8 | 9 | 7 | 3 | 6 | 5 | 4 | 1 | 3 | 1 | 4 | 4 | 5 | 5 |
| | 3 | 7 | 6 | 6 | 6 | 5 | 8 | 8 | 6 | 3 | 4 | 1 | 8 | 1 | 3 | 1 | 4 | 3 | 6 | 6 |
| I | 10 | 5 | 6 | 6 | 8 | 6 | 8 | 10 | 7 | 10 | 5 | 8 | 3 | 10 | 7 | 7 | 8 | 8 | 0 | 8 |
| | 4 | 7 | 7 | 6 | 7 | 6 | 5 | 7 | 7 | 6 | 4 | 7 | 5 | 8 | 9 | 8 | 9 | 8 | 6 | 6 |
| | 3 | 7 | 6 | 7 | 10 | 6 | 8 | 9 | 8 | 6 | 8 | 8 | 7 | 6 | 5 | 6 | 5 | 6 | 10 | 7 |
| | 7 | 8 | 6 | 7 | 8 | 8 | 9 | 8 | 6 | 6 | 5 | 7 | 8 | 6 | 3 | 6 | 6 | 7 | 7 | 7 |
| | 9 | 7 | 6 | 6 | 8 | 8 | 10 | 8 | 6 | 6 | 7 | 7 | 3 | 7 | 4 | 7 | 6 | 4 | 5 | 7 |

Table 3. Experts (A–I) give points to the top 5 ranked data by the proposed method (PM) and by the standard BM25 search method (BM), respectively

| | A | B | C | D | E | F | G | H | I |
|------|------|------|-------|------|------|-------|------|------|------|
| PM | 30.7 | 45.9 | 33.41 | 52.8 | 24.5 | 40.98 | 69.7 | 38.3 | 67.2 |
| BM25 | 21.6 | 75 | 30.49 | 65.5 | 35.7 | 33.92 | 64.9 | 37.8 | 86.1 |

Table 4. Comparative result between the proposed method (PM) and the standard BM25 using Equation (2). A–I are experts.

the collective intelligence. Web 2.0 emphasizes the human participations. Human judgment is quite often more accurate than the computer algorithm method for evaluating content values. For example, we can tell whether or not a given video has a car scene easily whereas computer cannot.

We have examined the human participations in video sharing sites. We consider YouTube as a sample site since it is the most popular video sharing site and has many nice features for verifying our approach. We have defined the participations to be social activities. Then, based on social activities, we have established a social network of users and videos in YouTube and have proposed an algorithm that computes user reputations and video reputations. We have shown that user reputation is quite related to the number of subscriptions and the number of uploaded contents. For the search usefulness of user reputation, we have comparative experiments between the proposed method and the standard BM25 method. From the results the proposed method is meaningful to find valuable video contents.

In future, we shall compute content score using the user reputation as a weight function. For example, if a user with high reputation gives a score or a comment, then the score or the comment would have more weight than a similar one by a user with low reputation. For this, we need to keep all the record of user activities and it is not an easy task. For this, we intend to set up a video sharing site.

Acknowledgments

We wish to thank the referee for the careful reading of the paper and many valuable suggestions.

Han was supported by the Basic Science Research Program through NRF funded by MEST (2010-0009168). Kim and Cha were supported by the IT R&D program of MKE/IITA 2008-S-024-01. Jeong-Won Cha is a corresponding author.

REFERENCES

- [1] AGARWAL, N.—LIU, H.—TANG, L.—YU, P. S.: Identifying the Intentional Bloggers in a Community. In Proceedings of the International Conference on Web Search and Web Data Mining, WSDM 2008, pp. 207–218, 2008.
- [2] BA, S.—PAVLOU, P.: Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. MIS Quarterly, Vol. 26, 2002, No. 3, pp. 243–268.

- [3] BENNETT, T. R.—BOOKER, J. M.—KELLER-McNULTY, S.—SINGPURWALLA, N. D.: Testing the Untestable: Reliability in the 21st century. *IEEE Transactions on Reliability*, Vol. 52, 2003, No. 1, pp. 118–124.
- [4] BIELIKOVÁ, M.—ANDREJKO, A.: Comparing Instances of Ontological Concepts for Personalized Recommendation in Large Information Spaces. *Computing and Informatics*, Vol. 28, 2009, No. 4, pp. 429–452.
- [5] BIELIKOVÁ, M.—BARLA, M.—TVAROŽEK, M.: Rule-Based User Characteristics Acquisition from Logs with Semantics for Personalized Web-Based Systems. *Computing and Informatics*, Vol. 28, 2009, No. 4, pp. 399–427.
- [6] BRIN, S.—PAGE, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, Vol. 30, 1998, No. 1-7, pp. 107–117.
- [7] DELLAROCAS, C.: Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior. In *EC'00: Proceedings of the 2nd ACM conference on Electronic commerce*, ACM 2000, pp. 150–157.
- [8] DELLAROCAS, C.: Analyzing the Economic Efficiency of Ebay-Like Online Reputation Reporting Mechanisms. In *EC'01: Proceedings of the 3rd ACM Conference on Electronic Commerce 2001*, pp. 171–179.
- [9] GLINER, J.—MORGAN, G.—HARMON, R.: Measurement Reliability. *Journal of the American Academy of Child & Adolescent Psychiatry*, Vol. 40, 2001, No. 4, pp. 486–488.
- [10] GYÖNGYI, Z.—GARCIA-MOLINA, H.—PEDERSEN, J.: Combatting Web Spam with Trustrank. In *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB) 2004*, pp. 576–587.
- [11] HAN, Y. S.—KIM, L.—CHA, J. W.: Evaluation of User Reputation on YouTube. In *HCI 12, 2009*, pp. 346–353, *Lecture Notes in Computer Science* No. 5621.
- [12] JONES, K. S.—WALKER, S.—ROBERTSON, S. E.: A Probabilistic Model of Information Retrieval: Development and Comparative Experiments. *Information Processing and Management*, Vol. 36, 2000, No. 6, pp. 779–808.
- [13] JØSANG, A.: The Beta Reputation System. In *Proceedings of the 15th Bled Electronic Commerce Conference 2002*.
- [14] KRISHNAN, V.—RAJ, R.: Web Spam Detection With Anti-Trust Rank. In *AIR-Web '06: Proceedings of the 2nd international workshop on Adversarial information retrieval on the web 2006*, pp. 37–40.
- [15] O'REILLY, T.: What is web 2.0 – Design patterns and business models for the next generation of software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, 2005.
- [16] RADOVANOVIĆ, M.—IVANOVIĆ, M.—BUDIMAC, Z.: Text Categorization and Sorting of Web Search Results. *Computing and Informatics*, Vol. 28 2009, No. 6, pp. 861–893.
- [17] RESNICK, P.—ZECKHAUSER, R.: Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. In M. R. Baye (Ed.): *The Economics of the Internet and E-Commerce*. Volume 11 of *Advances in Applied Microeconomics*, pp. 127–157, Elsevier Science 2002.

- [18] SABATER, J.—SIERRA, C.: Reputation and Social Network Analysis in Multi-Agent Systems. In AAMAS '02: Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems 2002, pp. 475–482.
- [19] STANDIFIRD, S. S.: Reputation and E-Commerce: eBay Auctions and the Asymmetrical Impact of Positive and Negative Ratings. *Journal of Management*, Vol. 27, 2001, No. 3, pp. 279–295.
- [20] ZHANG, Y.—ZHANG, L.—ZHANG, Y—LI, X.: Xrank: Learning More from Web User Behaviors. In CIT '06: Proceedings of the Sixth IEEE International Conference on Computer and Information Technology 2006, p. 36.



Yo-Sub HAN obtained his Ph.D. in computer science from HKUST in 2006. He worked as a researcher in KIST until 2009 and is currently an Assistant Professor at the department of Computer Science in Yonsei University. His research interests include formal language theory, algorithm design and information retrieval.



Laehyun KIM is a principal research scientist in Center for Bionics at Korea Institute of Science and Technology. His research interests include haptics, computer graphics, and virtual reality. He received his BS from Hanyang University in material engineering, MS from Yonsei University in Computer Science, and Ph.D. in computer science from the University of Southern California.



Jeong-Won CHA received his B.Sc. degree in computer engineering from Soongsil University in 1996. He received M.Sc. and Ph.D. in computer engineering from POSTECH in 1999 and 2002, respectively. He was a post-doctor in USC/ISI from March 2002 to February 2003. He joined Changwon National University in 2004 and is now an Associate Professor. His research interests currently include natural language processing and information retrieval.