Ranking User-Created Contents by Search User's Inclination in Online Communities

Hyoseop Shin
Dept. of Advanced Technology Fusion
School of Internet & Multimedia Eng.
Konkuk University, Seoul, Korea
+82-2-2049-6117
hsshin@konkuk.ac.kr

Jeehoon Lee
Dept. of Advanced Technology Fusion
School of Internet & Multimedia Eng.
Konkuk University, Seoul, Korea
+82-10-3335-0641
jeehoon.lee01@gmail.com

ABSTRACT
Searching posts effectively has become an important issue in large-scale online communities. Especially, if search users have different inclinations when they search posts, they have different kinds of posts in their minds. To address this problem, in this paper, we propose a scheme of ranking posts based on search users' inclination. User ranking score is employed to capture posts that are relevant to a specific user inclination. Specifically, we present a scheme to rank posts in terms of user expertise and popularity. Experimental results show that different user inclinations can produce quite different search results and the proposed scheme achieves about 70 % accuracy.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Information Filtering

General Terms
Algorithms

Keywords
online community, search user inclination, user ranking, post ranking, expertise, popularity

1. Introduction
Online communities on the web are so popular that large-scale ones like YouTube often contain up to millions of posts created by community users. Thus, searching posts effectively has become an important issue. Especially, if search users have different inclinations when they search posts, they have different kinds of posts in their minds. For example, popular posts must be ranked high when search users look for popular contents, meanwhile high-quality posts should be ranked high when they want contents of high expertise. In this paper, we propose a scheme of ranking posts based on search users' inclination.

2. User Inclination-based Post Ranking
A post's score is computed by the combination of semantic similarity score and user inclination-based score as follows:

$$\mathrm{PostScore}(query, post) = \alpha * \mathrm{SemanticSimilarity}(query, post) + (1-\alpha) * \mathrm{UserInclinationPostScore}(post)\,, \quad (0 \le \alpha \le 1)$$

The semantic similarity score of a post against a search query is computed by a TF*IDF-based document weighting scheme. The user inclination-based score of a post is obtained by the summation of the ranking scores of the users who have given any contribution to the content:

$$\mathrm{UserInclinationPostScore}(\mathrm{p}) = \sum_{\mathrm{u} \in \mathrm{U_p}} \mathrm{UserRankingScore}(\mathrm{u})$$

, where $\mathrm{U_p}$ consists of the users that have contributed to the post, p. Here, contribution means post-related user activities such as posting the post, or comment, read, and recommendation on the post. The method to compute user ranking scores should be dependent on the specific user inclinations. In this paper, we introduce two user inclinations: expertise and popularity. Figure 1 represents the framework of the scheme. In off-line step, user ranking scores[1] on specific user inclinations are computed a priori, which is followed by user score transform process. When a user query comes in, user inclination-based post scores are combined with semantic similarity scores to produce the rank results.

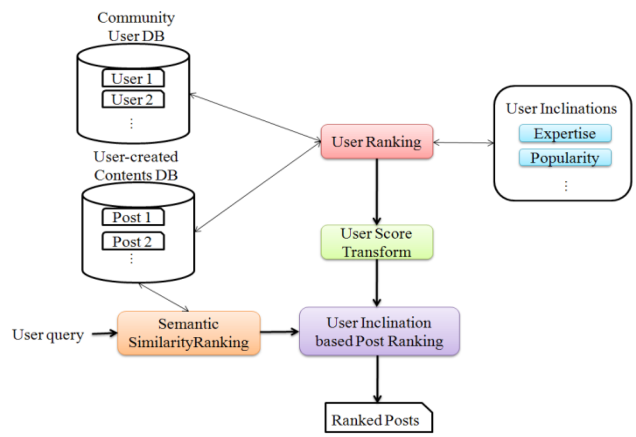


Figure 1. Framework of User Inclination-based Post Ranking

3. User Expertise and Popularity
User-created contents like photos or videos can be measured in two contrary aspects: expertise and popularity. While some users want to see popular photos that many other users have seen, others may want to see photos that are regarded to have high expertise. In this paper, we present a scheme to model these two aspects as the comparative user inclinations.

If a post contains high-quality or expert-level contents, the post is likely to receive more responses from expert users than from non-expert users. In contrast, if a post contains low-quality or novice-level contents, expert users may not respond the post at all. It can be said that the expertise of a post can be determined by the expertise of the commentators for the post. In other words, the commentators' expertise can affect the expertise of the poster.

From this inference, we design the user expertise feature: *Expertise Reference*(ER). The ER of a user *u* is affected by the ER values of all the users *v* who have commented on the posts written by the user *u*.

Let $|\mathrm{A_u}|$ be the number of posts written by the user *u* and $|\mathrm{C_v}|$ be the number of comments written by the user *v*, and $|\mathrm{C_{A_u,v}}|$ be the number of comments written by the user *v* for the posts Au, which is written by the user *u*. Then, the expertise reference, ER(u) of the user *u* is formulated as follows:

WWW 2009, April 20-24, 2009, Madrid, Spain.
ACM 978-1-60558-487-4/09/04.

$$ER(u) = d * \sum_{v} \frac{|C_{A_u,v}|}{|C_v|} ER(v) + (1 - d)$$

, where $\sum_u |C_{A_u,v}| = |C_v|$ and d is the damping factor.
Meanwhile, popularity could be thought as the opposite concept of expertise. However, some posts with high expertise can be also popular. Popularity had better be translated as the extent of inclinations that normal users like. To reflect this stream, we obtain the popularity scores of users by applying transform function such as normal distribution for user ranks in accordance with the user expertise scores explained above.

4. User Ranking Score Transform

Each user's score obtained from user ranking on a specific user inclination criterion can be additionally tailored by user score transform functions to meet different user inclination criteria. For example, given user ranking of expertise, if a search user wants to see low-browed but popular contents, a transform function will effect to make comparatively high user scores into lower ones and low scores into higher ones. In some cases, the transformation from user scores into user ranks may be required. Furthermore, user ranks can be passed into probability density functions to produce biased user scores. For example, normal distribution and exponential distribution can give more weight to average-ranked users and high-ranked users, respectively.

5. Experimental Results

We conduct the experiments with the data sets from a large-scale web community called "SLR Club (http://www.slrclub.com)" in which users can post photos and also give comments on other users' photos. The data set contains 129,941 posts and 1,191,716 comments by 114,308 users for about one and half years. In ranking posts in term of expertise, we use the combination of ER and identity transform. As for popularity, we combine ER and normal distribution probability density transform on ER ranks.

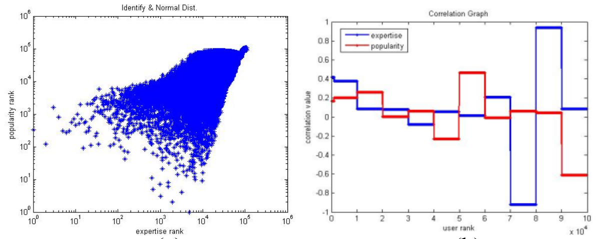
Figure 2 demonstrates some high-ranked photos on expertise and popularity searched by the proposed scheme. In each figure 2-(a) and 2-(b), upper four pictures are results of no search keyword, and lower four are results of search keyword "sea". Photos of expertise appear to be quite relevant to high-level skills on photography; meanwhile, popular photos seem to be related to interestingness of stuff such as young woman, unusual situation, rare animals, and so on.

(a) Photos of Highest Expertise

(b) Photos of Highest Popularity

Figure 2. Retrieved Photos with the highest expertise and with the highest popularity.

We analyze the correlation between expertise and popularity of posts. As can be seen in figure 3-(a), the correlation is minimal among high-ranked expertise or popularity posts, and it gets higher in low-ranked posts. The x-axis represents the post ranks in order of expertise and y-axis represents popularity ranks.



(a) (b)

Figure 3. Correlation between expertise and popularity of posts. (a): post distribution on expertise and popularity dimensions. (b): Pearson's correlation fluctuation over user ranks.

Figure 3-(b) also shows the trend that the correlation fluctuation gets larger as the post rank increases.

Finally, to evaluate the performance of judgment on expertise and popularity of the proposed scheme, we let four persons to judge 253 photos whether each photo has more expertise or more popularity. By the proposed scheme, among the 253 pictures, 128 pictures are categorized as high expertise ones, and 125 pictures as high popularity ones. Table 2 summarizes the user study results. As for precision, photos within the popularity category get higher score (71.2%), meanwhile photos within the expertise category get higher score (69.2%) in recall. The overall F1 score is about 66%~68%. The demonstration is available at our web site[2].

Table 1. User Study Results

	Expertise	Popularity
# of pictures	117	136
# of predictions	128	125
# of detects	81	89
# of false alarms	47	36
# of false dismissals	36	47
Precision (%)	63.3 %	**71.2 %**
Recall (%)	**69.2 %**	65.4 %
F1 score (%)	66.1 %	68.2 %

6. Conclusions

In this paper, we presented a scheme to effectively find user-created contents based on search users' inclinations in large-scale online communities. Experimental results showed that user inclination is very influential in ranking posts diversely. Also, we proposed that user inclinations could be captured by user ranking mechanism. As a case study, we presented user expertise and popularity. It should be noted that our approach of using user inclination in searching online community contents is different from the user-profile based personalized web search approaches like the recent publication[3]. Future work includes further investigation on possible user inclinations.

7. Acknowledgements

This work was partially supported by Seoul R&BD Program in Korea and the Korea Science and Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund)(KRF-2008-313-D00864).

8. References

[1] Hyoseop Shin, Zhiwei Xu, and Eun Yi Kim, "Discovering and Browsing of Power Users by Social Relationship Analysis in Large-scale Online Communities," Proc. of 2008 IEEE/WIC/ACM conf. on Web Intelligence, Sydney, Australia, Dec. 2008.

[2] Expertise/Popularity-based Community Search Demo, "http://webi.konkuk.ac.kr/ep_rank".

[3] J. Luxenburger, S. Elbassuoni, and G. Weikum, "Matching task profiles and user needs in personalized web search," Proc. of 2008 ACM CIKM conference, pages 689-698, Napa Valley, USA, Sep. 2008.