

FolkPopularityRank: Tag Recommendation for Enhancing Social Popularity using Text Tags in Content Sharing Services

Toshihiko Yamasaki, Jiani Hu, Shumpei Sano, and Kiyoharu Aizawa

Department of Information and Communication Engineering, The University of Tokyo
yamasaki@ay-lab.org

Abstract

In this study, we address two emerging yet challenging problems in social media: (1) scoring the text tags in terms of the influence to the numbers of views, comments, and favorite ratings of images and videos on content sharing services, and (2) recommending additional tags to increase such popularity-related numbers. For these purposes, we present the FolkPopularityRank algorithm, which can score text tags based on their ability to influence the popularity-related numbers. The FolkPopularityRank algorithm is inspired by the PageRank and FolkRank algorithms but the scores of the tags are calculated not only by the co-occurrence of the tags but also by considering the popularity-related numbers of the content. To the best of our knowledge, this is the first attempt to recommending tags that can enhance popularity attributes of social media. We conducted extensive experiments with about 1,000 images. We uploaded the photos with the recommended tags along with the original tags to Flickr as a real test, and obtained very promising results.

1 Introduction

In image and video sharing services such as Flickr¹, Instagram², etc., the numbers of views, comments, and favorites (likes) are shown to represent its popularity. Hereafter, we refer to such numbers as “social popularity scores” or “social popularity,” aligned with the definition in [Yamasaki *et al.*2014]. It is observed that users often care about by their social popularity scores in social networks. Users even sometimes do self-promotion to gain positive sentiments (e.g., asking their friends to show “favorite” or “like” for them), not to mention the companies which can leverage social media services for promotion. Therefore, in social media networks, it has been a critical issue for both individuals and enterprises to obtain high social-popularity scores since such scores directly reflect how much attention has been paid to certain contents.

¹<https://www.flickr.com/>

²<https://instagram.com/>

Though one might think that quality of the image and video content is one of the most important factors in receiving higher social popularity scores in social networks, it has been shown that the image and video feature does not contribute much in predicting social popularity scores [Yamaguchi *et al.*2014, Gelli *et al.*2015, Hu *et al.*2016]. Text tags, on the other hand, play a very important role, because users typically locate content via text searches. Therefore, predicting social popularity scores using text tags along with other features has been studied in [Pedro and Siersdorfer2009, van Zwol *et al.*2010, Yamasaki *et al.*2014]. The review on such prediction of social popularity scores is summarized in Sec. 1.1

It is worthwhile to question whether users have sufficient knowledge to annotate their content with tags that are capable of boosting social popularity scores. In many cases, they lack this knowledge. Users often do not know which tags are more influential, and they cannot predict social popularity scores based only on tags. We conducted a preliminary experiment by uploading three identical photos to three different new accounts in Flickr, as shown in Fig. 1. The first uploaded photo included the original tags (*lisbon, portugal, homeless, smoking, bw, thecontinuum, 1on1people, and myphotobook*); the second included the top two tags in terms of tag frequency ranking (*bw and portugal*), and the third included the top two recommended tags from our system (*bw and smoking*). Surprisingly, four days after the upload, the numbers of views were 29, 8, and 30, respectively. The image with the recommended tags earned 3.8 times more views than the tag set #2. This experiment clearly shows that even though the images are identical, their social popularity scores are strongly affected by the tags.

In this paper, we propose a new algorithm called FolkPopularityRank (or FP-Rank). FP-Rank is inspired by PageRank [Brin and Page1998, Page *et al.*1999], the well-known web page ranking algorithm, and FolkRank [Hotho *et al.*2006b, Hotho *et al.*2006a], PageRank’s folksonomy-based descendant that also considers users, resources, and tags. The contributions of this paper is that FP-Rank can recommend tags that enhance social popularity scores. The recommendation is done based on existing tags, and therefore the recommended tags are related to the content of the image/video. Although numerous tag recommendation systems have already been proposed (please refer to [Li *et al.*2015b] for reviews), they are focused only on tags that are semantically correct for

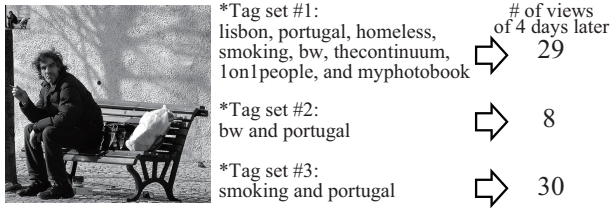


Figure 1: Preliminary results demonstrating the importance of proper tagging

describing the content. To the best of our knowledge, tag recommendations designed to increase social popularity scores is firstly presented.

1.1 Related Works

PageRank and FolkRank

Our algorithm is inspired by the PageRank [Brin and Page1998, Page *et al.*1999] and FolkRank [Hotho *et al.*2006b, Hotho *et al.*2006a] algorithms. PageRank was developed to rank web pages. The basic concept of PageRank consists of three assumptions: (1) web pages linked to by many other pages are important, (2) web pages linked to by important pages are also important, and (3) if a page contains links to many other pages, the importance of being linked by the page is low. Thus, PageRank regards web pages as a directed graph, in which edge weights are the importance score of the source page divided by the number of outgoing links. For computational efficiency, the scores \mathbf{r}_P are calculated iteratively by the following equation:

$$\mathbf{r}_P = \alpha A_P \mathbf{r}_P + (1 - \alpha) \mathbf{p}, \quad (1)$$

where \mathbf{r}_P is the importance score vector and A_P is the row-stochastic version of the adjacency matrix of the directed graph $G_P = (V_P, E_P)$. α is a dumping factor that can consider random jumps to other pages, and \mathbf{p} is the random surfer component.

FolkRank is an extension of PageRank, and considers users, resources, and tags. FolkRank is constructed with an undirected tripartite graph $G_F = (V_F, E_F)$, where $V_F = U \cup R \cup T$. U , R , and T represent users, resources, and tags, respectively. Edge weights represent all the co-occurrences of tags and users, users and resources, and resources and tags: $E_F = \{\{t, u\}, \{u, r\}, \{r, t\}\}$. The scores \mathbf{r}_F can be calculated by the following equation.

$$\mathbf{r}_F = \alpha A_F \mathbf{r}_F + (1 - \alpha) \mathbf{p}, \quad (2)$$

where A_F is the row-stochastic version of the adjacency matrix G_F . It is assumed that resources tagged with important tags (and annotated by important users) are important; this assumption applies symmetrically to users, resources, and tags.

As already noted in [Brin and Page1998], the random surfer component \mathbf{p} can be used as a personalized ranking. Further, tag recommendations based on already annotated tags are calculated by the FolkRank algorithm, using the following equation.

$$\mathbf{w}_F = \mathbf{r}_F^{(d=1)} - \mathbf{r}_F^{(d=0)}. \quad (3)$$

However, the recommended tags are only co-occurrence-based, which does not consider the effect to the social popularity at all.

Tag Ranking

One of the main purposes of tag ranking in social media is to extract influential tags for (1) classification of the content or (2) regression of the scores, in terms of attractiveness or popularity. In [Pedro and Siersdorfer2009], top terms for attractive and unattractive photos in Flickr were extracted according to mutual information values. It was shown that a regression model using the extracted tags can achieve a correlation value of 0.36 for the predicted and actual number of favorites in Flickr. In addition, visual features such as brightness and contrast were introduced. The combination of text and visual features achieved a correlation value of 0.48. In [Yamasaki *et al.*2014], a tf-idf-like tag ranking was proposed using the tag frequency and weights for the tags, which were obtained from a regression model.

Social Popularity Prediction

Predicting how much popularity the content would get is another important research topic [van Zwol *et al.*2010, Khosla *et al.*2014, McParlane *et al.*2014, Totti *et al.*2014, Cappallo *et al.*2015, Gelli *et al.*2015, Chen *et al.*2016b, Wu *et al.*2016]. In [van Zwol *et al.*2010], social connections, in addition to textual and visual information, were also utilized to predict which content would earn a higher number of favorites. [Khosla *et al.*2014] showed that social context is very important to predict the social popularity and they achieved the correlation value of 0.81. [Totti *et al.*2014] investigated the impact of the visual attributes and showed that visual properties have low predictive power compared that of social cues. In [McParlane *et al.*2014], an image's context, visual appearance, and user context were considered. [Gelli *et al.*2015] introduced visual sentiment features in addition to the above mentioned features. In most papers, context data such as tags were most dominant.

The disadvantage of using visual features is their computational complexity. Because billions of photos and videos exist on content sharing services, extracting visual features for all of them is not practical. It has also been demonstrated that visual features do not contribute much in some content sharing services [Yamaguchi *et al.*2014, Gelli *et al.*2015, Hu *et al.*2016]. The use of social connections suffers from the dynamic nature of the connections themselves. In addition, social connections cannot be used to analyze the influence of new users who do not yet have connections; this situation is known as the cold start problem. It is also difficult to control the social connections: even if the number of social connections with influential persons is important for higher social popularity, for instance, it is not easy to add such connections.

Tag Recommendation and Refinement

Tag recommendation and refinement in social network help users to annotate more tags with less effort, and to consolidate vocabulary across users.

One of the most prominent and frequently used recommendation techniques is collaborative filtering (CF). Because

there are so many CF-based methods, please refer to the review papers [Su and Khoshgoftaar2009, Lops *et al.*2013] for details.

In Tagcoor [Sigurbjörnsson and van Zwol2008], candidate tags were extracted by evaluating their co-occurrences; they were sorted using a combination of three measures: stability, descriptiveness, and user-based tag frequency ranking.

FolkRank [Hotho *et al.*2006b, Hotho *et al.*2006a] can also be used for tag recommendations, as summarized in Sec. 1.1. Numerous techniques have been proposed to improve FolkRank. [Jschke *et al.*2007] compared several approaches, including adapted PageRank, FolkRank, CFs, and co-occurrence-based rankings; they demonstrated that FolkRank provides the best performance. In [Si *et al.*2009], the most popular tags and FolkRank were combined. [Gem-mell *et al.*2009] and [Zhang *et al.*2009] separately proposed hybrid recommendation models in which tag rankings were calculated by the weighted sum of FolkRank scores and item-based CF scores. [Landia *et al.*2012] proposed Content-FolkRank by combining FolkRank with textual content to solve the cold start problem.

A comprehensive survey can be found in [Li *et al.*2015b]. About 60 approaches are summarized and 11 of them are re-implemented by the authors and evaluated using the same dataset. We would like to emphasize that the authors in [Li *et al.*2015b] compare the performance only in terms of average precision (AP). In other words, the tag recommendation and retrieval is regarded as an information retrieval problem.

Tag Generation and Scene Description

Object-recognition-based sentence generation systems [Kulkarni *et al.*2013, Han *et al.*2014, Chen *et al.*2016a] can be regarded as automatic tag generators. Although the descriptions are semantically correct, the sentences and descriptions are nothing more than objective facts. Therefore, these systems are useful for content retrieval by text search; however, they do not consider factors that would increase social popularity when generating descriptions.

Attractiveness/Emotion Prediction

Human emotion-based attractiveness of content has been investigated over the past decade. Sentiment classification and affective analysis have been conducted for texts [Pang *et al.*2002, Pang and Lee2008], images [Lu *et al.*2014, Li *et al.*2015a, Peng *et al.*2015, Zhao *et al.*2016], and videos [Irie *et al.*2010]. Aesthetic analysis of the content using image features [Bhattacharya *et al.*2010, Murray *et al.*2012, Zhang *et al.*2014, Lu *et al.*2014, Aydin *et al.*2015, Park and Zhang2015] has also been a frequently discussed topic in the community. However, we would like to point out that these papers did not discuss how such emotion-based attractiveness would affect the social popularity.

2 Folk Popularity Rank

2.1 Tag Ranking Algorithm

We would like to emphasize that conventional tag ranking and recommendation systems in social media are designed to add semantically proper tags. Namely, some of the tags in

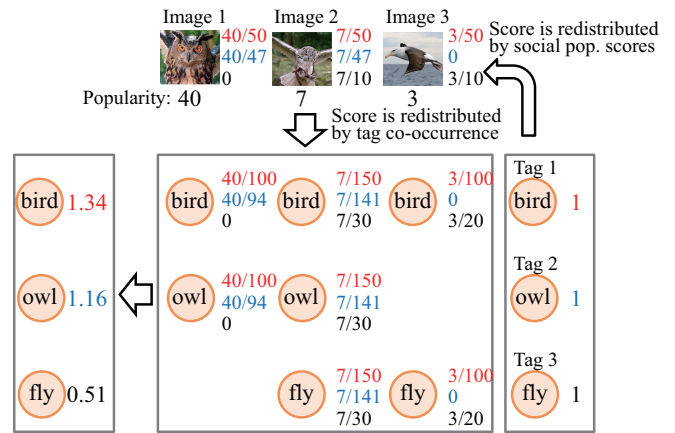


Figure 2: A toy example to show how FP-Rank calculates scores for the tags (for the first iteration).

the original content were artificially deleted, and the recommender systems attempted to predict as many deleted tags as possible while suppressing the number of unrelated tags.

We propose an algorithm that can extract tags with a higher level of influence on social popularity scores. As a result, we can recommend tags that enhance social popularity. It is important to note that visual and social features can also be employed, as in [Pedro and Siersdorfer2009, van Zwol *et al.*2010, Yamaguchi *et al.*2014]. However, this paper focuses only on tags, to emphasize the importance of tags.

The assumption here is that (1) tags used for images and videos with high social popularity scores are important, (2) the tags co-occurring with such important tags are also important, and (3) the tags become less important when annotated with many other tags. The equation for our FP-Rank scores, s , is defined as follows.

$$\mathbf{r}_{FP} = \alpha A_{FP} \mathbf{r}_{FP} + (1 - \alpha) \mathbf{p}, \quad (4)$$

$$\mathbf{s} = \mathbf{r}_{FP} - \mathbf{r}_F^{tag}, \quad (5)$$

where \mathbf{r}_F^{tag} is the tag-only FolkRank score.

There are two fundamental differences from PageRank and FolkRank. The first involves the design of the matrix A_{FP} . A_{FP} is a square matrix whose size is $T \times T$, where T is the number of unique tags in the service. The matrix A_{FP} can be interpreted as a combination of two matrices by introducing the tag-content matrices:

$$A_{FP} = B_w \times B_t^T. \quad (6)$$

Here, B_w and B_t are $T \times C$ row-stochastic and column-stochastic matrices, respectively, where C is the number of the images and videos in the service. The i th row vector in B_w is a set of social popularity scores for the content tagged with the tag i , normalized by the sum of the scores in the row. The j th column vector in B_t is a set of tag usage flags normalized by the total number of tags attached to the content c . As a result, the component of the matrix A_{FP} , a_{ij} is calculated as follows.

$$a_{ij} = \sum_{d \in \mathbf{D}} \frac{u(d)_i}{(\# \text{ of tags attached to content } d)}, \quad (7)$$

where d is the index of the image or video content that contains both tags i and j , \mathbf{D} is the training dataset, and $u(d)_i$ is the weight for the tag i annotated to the content d . In FolkRank, $u(d)_i$ is simply calculated by

$$u(d)_i = \frac{\text{(current score for tag } i)}{\text{(\# of images/videos annotated with tag } i)}. \quad (8)$$

In contrast, FP-Rank calculates $u(d)_i$ by considering the content’s social popularity scores:

$$u(d)_i = \frac{\text{(current score for tag } i) \times \text{(social pop. score of content } d) + k}{\sum \text{(social pop. score of content with tag } i)}, \quad (9)$$

where k is a parameter to prevent $u(d)_i$ from being 0 and $k = 1$ in our study.

Using this procedure, tag scores are redistributed to other tags via the image/video, through weighting determined by social popularity. In other words, FP-Rank considers not only the number of links (co-occurrences) between tags, but also the popularity of the content to which the tag is annotated.

The concept of the procedures in eqs. (7) and (9) is visually explained by a toy example in Fig. 2. Let us assume that there are only three images and only three unique tags in the service. The initial importance scores for the tags are all set to 1. Image #1 is annotated with *bird* and *owl*, image #2 with *bird*, *owl*, and *fly*, and image #3 with *bird* and *fly*. The images’ social popularity scores (which include the numbers of views, comments, or favorites) are 40, 7, and 3, respectively. The tag scores are distributed to the images by considering the social popularity scores. They are collected back to the tags by considering the number of co-occurring tags. Then, the tag scores are updated by summing the scores of the tags. In this case, B_w and B_t are defined as follows.

$$B_w = \begin{pmatrix} 40 & 7 & 3 \\ 50 & 50 & 50 \\ 40 & 47 & 3 \\ 0 & 10 & 10 \end{pmatrix}, B_t = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{2} \end{pmatrix}. \quad (10)$$

The second difference from PageRank and FolkRank is that the FP-Rank score is given by eq. (5), not by eq. (4). In \mathbf{r}_{FP} in eq. (4), both co-occurrence-based scores and popularity-based scores are inevitably mixed. Therefore, by subtracting by the tag-only FolkRank score \mathbf{r}_F^{tag} , only the popularity-based scores are obtained.

The iteration in eq. (4) is repeated until it converges. The related literature [Page *et al.* 1999] points out that eq. (1) converges after approximately 50 iterations, but 10 iterations are sufficient for practical ranking systems. We confirmed that this is also the case with our FP-Rank algorithm.

Different from the original FolkRank, resource factors need not be considered because the social popularity score is given instead. When users are included in the FP-rank model, it is possible to conduct user-aware recommendations.

2.2 Tag Recommendation

Similar to FolkRank, tags can be recommended by the following equation.

$$\mathbf{w}_{FP} = \mathbf{s}^{(d=1)} - \mathbf{s}^{(d=0)}, \quad (11)$$

Table 1: Overview of training and test sets.

Data Set	Training	Test
Total Number of Images	60,000	947
Number of Unique Tags	254,734	17,387
Average Number of Views	13,139.5	221.4
Average Number of Tags	37.1	30.2

where \mathbf{w}_{FP} is a weight vector with one entry for each tag. The random surfer component for the already existing tags is set to 1, and the others are set to $0 - 1$; the eq. (4) is iterated until convergence. Setting the random surfer component \mathbf{p} in this manner causes tags co-occurring with the already existing tags to be extracted. If the random surfer components for the unused tags are 0, there is no chance of selecting random tags. In addition, as explained in Sec. 2.1, the tag scores are redistributed by considering the contents’ social popularity. Thus, the final tag scores correspond to the influence of the tags’ social popularity scores.

3 Experimental Results

The experiments were performed on the YFCC100M dataset [Thomee *et al.* 2015]. We randomly selected 60,000 images that have more than 20 tags and over 5000 views of the whole YFCC100m dataset for training and randomly selected 947 images annotated with over 20 tags but different number of views for testing (originally it was 1,000, but users dynamically updated/deleted their content and we could download only 947 images). Table 1 gives an overview on the training and testing sets. We can observe that the training set contains 250K unique tags and the images in the training set are really popular in Flickr, which can be good resources for recommendation.

3.1 Tag Recommendation

We first describe the procedure of tag recommendation. We evaluated a tag recommender based on the proposed FP-Rank and compared with four other recommendation methods: (1) Doing nothing (using only the original tags), (2) Tag-coor [Sigurbjörnsson and van Zwol2008], (3) Collaborative Filtering (CF) [Su and Khoshgoftaar2009, Lops *et al.* 2013], and (4) CF with DF-W [Yamasaki *et al.* 2014].

We added 10 tags for each image using the five recommendation methods respectively. An example of the original and the recommended tags are shown in Table 2. Changing the number of recommended tags is our future plan.

We evaluated the five methods in two ways. First, we uploaded a set of images and tags to Flickr and tracked the number of views. Second, the quality of the tag sets were evaluated humans in a crowd working service.

3.2 Uploading to Flickr

Online experiments is one of the most convincing approaches to evaluate the efficiency in popularity boosting. We repeated the following procedure five times: creating a new Flickr account, uploading images with tags that were recommended by one of the five approaches along with the original tags, checking the number of views for 10 days, and deleting all the files

Table 2: Example of recommendation results.

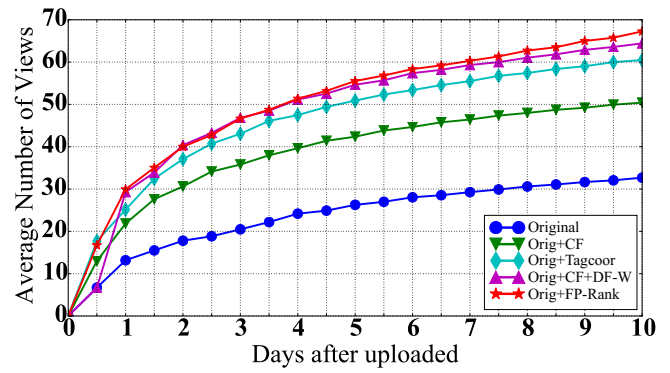


(b) Original tags and Recommendations

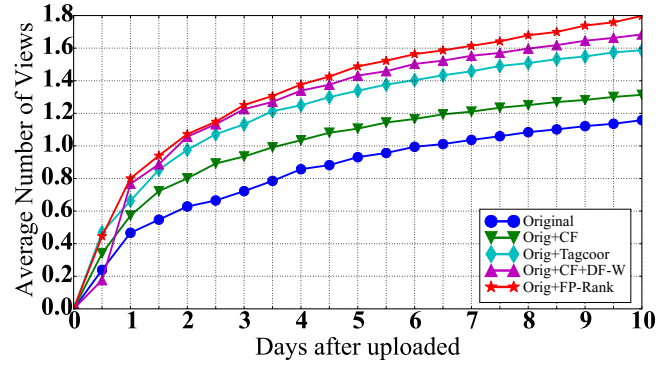
Approach	Tags
Original tags	gravestones, headstones, leaves, sigma, shadows, leicester-shire, park, pdeee454, saint wistans, sunlight, sunshine, tombstones, trees, wistow, 17-70mm, 450d, canon, church, grass, kilby
Tagcoor	light, nature, sky, leaf, landscape, green, hdr, clouds, blue, sun
CF	nature, beach, water, australia, sky, sea, landscape, flowers, photography, clouds
CF+DF-W	landscape, beach, water, nature, girls, skyline, oregon, clouds, australia, sea
FP-Rank	hdr, light, nature, sky, landscape, sun, clouds, blue, green, water

and the account. To avoid the situation where the search engine of Flickr returns a series of identical images at the same time during the retrieval process, only one approach is evaluated at one time. A new account is generated each time in order to eliminate the other possible effects such as the account being “liked.” Therefore, social connection-based approach is not valid in this case. In this way, the subsequent experiments will not be disturbed by previous results, thus it ensures that the experiments are done independently. We record the number of views of all the images every 12 hours (this did not affect the numbers of views) and the whole online experiment lasted from Nov. 2016 to Jan. 2017.

The results are shown in Figure 3. FP-Rank performed the best not only by the number of views per image but also that per image and the number of tags. The average number of views is more than 2 times higher than using only the original tags. This proves that FP-Rank makes better recommendations with a higher level of influence on popularity boosting over the other three tag recommendation methods. The popularity of the original tag sets was the worst, which shows that the ordinary people are not aware of the importance of tag selection. And CF with DF-W also worked well, only a little worse than FP-Rank, and significantly better than CF. We conducted the paired Student’s t-test between CF+DF-W and our FP-Rank in terms of the number of views and $p = 0.015 < 0.05$. Although the results using FolkRank is not presented, its performance is the same as r_F^{tag} because tags are annotated only by the owner of the content in Flickr.



(a)



(b)

Figure 3: Results of online evaluation: average number of views. (a) Per image, (b) per both image and tag.

3.3 Tag Quality Assessment

One may doubt that it is possible to assign tags which have a high level of popularity influence but totally unrelated to the image, which will lead to tag spam. In order to eliminate this confusion and prove that our recommendation methods do recommend proper tags based on the content of web objects, we conduct a quality assessment of tags obtained by different recommendation algorithms.

We conducted our assessment on Amazon’s Mechanical Turk [Buhrmester *et al.*2011]. We asked workers to rate a tag on a five-point scale, where 5 stars for good quality and 1 star for bad quality. The content of an image and necessary descriptions are provided for reference, such as title, owner, date taken and etc. Figure 4 shows the design of the assessment task. We assigned five different workers, who has a approval rate of 85 % or higher, to rate each image-tag pair. The whole assessment took about three weeks in Jan. 2017 and more than 1,500 workers contributed. We carefully checked all the answers and rejected answers that were uncompleted or of poor quality.

The rating scores of tags recommended by different methods are summarized in Table 3. Original tags and those recommended by Tagcoor are rated the highest, and the FP-Rank is rated a little lower (by 0.1 point). The results show that FP-Rank recommend reasonably good tags that can properly describe the images. On the other hand, the average rating of CF-based methods are the worst among all the methods.

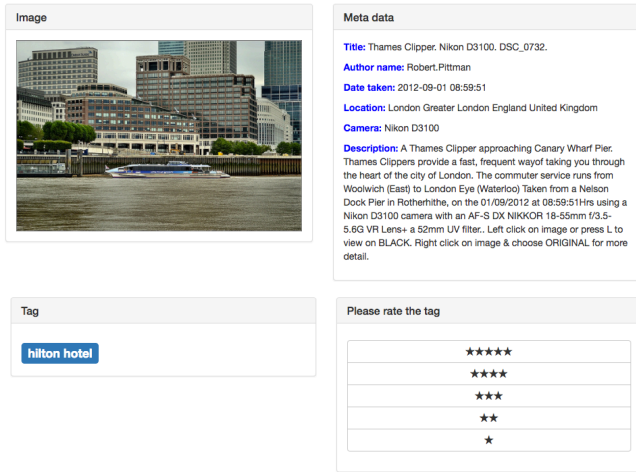


Figure 4: Sample design of tag quality assessment task.

Table 3: Results of tag quality assessment.

Tag Lists	# of tags	Ave. Score	Std. dev.
Original	26,958	2.9	1.3
CF	9,470	2.7	1.3
Tagcoor	9,470	2.9	1.3
CF+DF-W	9,470	2.7	1.3
FP-Rank	9,470	2.8	1.3

4 Conclusions

In this paper, a FolkPopularityRank algorithm was proposed to rank tags in content sharing services according to the order of their influence on social popularity scores (i.e., the numbers of views, comments, favorites, and other rankings). This concept is markedly different from conventional tag ranking and recommendation systems, which only recommend related tags without considering their impact on social popularity. We conducted a real-life experiment by uploading 947 image sets to Flickr, and showed a potential that the tags recommended by our system could efficiently increase social popularity while maintaining appropriateness to describe the content.

Acknowledgments

This work was partially supported by the Grants-in-Aid for Scientific Research (no. 26700008) from JSPS, JST-CREST (JPMJCR1686), and Microsoft IJARC core13. J. Hu is now with Microsoft Development, Co., Ltd. and S. Sano is now with Yahoo Japan Corporation.

References

[Aydin *et al.*, 2015] Tunc Ozan Aydin, Aljoscha Smolic, and Markus Gross. Automated aesthetic analysis of photographic images. *IEEE TVCG*, 21(1):31–42, Jan 2015.

[Bhattacharya *et al.*, 2010] Subhabrata Bhattacharya, Rahul Sukthankar, and Mubarak Shah. A framework for photo-

quality assessment and enhancement based on visual aesthetics. In *ACMMM*, pages 271–280, 2010.

[Brin and Page, 1998] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30:107–117, 1998.

[Buhrmester *et al.*, 2011] Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. Amazon’s mechanical turk a new source of inexpensive, yet high-quality, data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[Cappallo *et al.*, 2015] Spencer Cappallo, Thomas Mensink, and Cees G.M. Snoek. Latent factors of visual popularity prediction. In *ICMR*, pages 195–202, 2015.

[Chen *et al.*, 2016a] Jia Chen, Qin Jin, and Yifan Xiong. Semantic image profiling for historic events: Linking images to phrases. In *ACMMM*, pages 1028–1037, 2016.

[Chen *et al.*, 2016b] Jingyuan Chen, Xuemeng Song, Liqiang Nie, Xiang Wang, Hanwang Zhang, and Tat-Seng Chua. Micro tells macro: Predicting the popularity of micro-videos via a transductive model. In *ACMMM*, pages 898–907, 2016.

[Gelli *et al.*, 2015] Francesco Gelli, Tiberio Uricchio, Marco Bertini, Alberto Del Bimbo, and Shih-Fu Chang. Image popularity prediction in social media using sentiment and context features. In *ACMMM*, pages 907–910, 2015.

[Gemmell *et al.*, 2009] Jonathan Gemmell, Thomas Schimoler, Maryam Ramezani, Laura Christiansen, and Bamshad Mobasher. Improving folkrank with item-based collaborative filtering. In *ACM RecSys Workshop on RSWEB*, pages 17–24, 2009.

[Han *et al.*, 2014] Yahong Han, Xingxing Wei, Xiaochun Cao, Yi Yang, and Xiaofang Zhou. Augmenting image descriptions using structured prediction output. *IEEE TMM*, 16(6):1665–1676, 2014.

[Hotho *et al.*, 2006a] A. Hotho, R. Jäschke, C. Schmitz, G. Stumme, and K.-D. Althoff. FolkRank: A ranking algorithm for folksonomies. In *LWA*, volume 1, pages 111–114, 2006.

[Hotho *et al.*, 2006b] Andreas Hotho, Robert Jschke, Christoph Schmitz, and Gerd Stumme. Information retrieval in folksonomies: Search and ranking. In *ESWC*, pages 411–426, 2006.

[Hu *et al.*, 2016] Jiani Hu, Toshihiko Yamasaki, and Kiyoharu Aizawa. Multimodal learning for image popularity prediction on social media. In *ICCE-TW*, pages 264–265, 2016.

[Irie *et al.*, 2010] Go Irie, Takashi Satou, Akira Kojima, Toshihiko Yamasaki, and Kiyoharu Aizawa. Affective audio-visual words and latent topic driving model for realizing movie affective scene classification. *IEEE TMM*, 12(6):523–535, 2010.

[Jschke *et al.*, 2007] Robert Jschke, Leandro Marinho, Andreas Hotho, Lars Schmidt-Thieme, and Gerd Stumme. Tag recommendations in folksonomies. In *ECML PKDD*, pages 506–514, 2007.

- [Khosla *et al.*, 2014] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. What makes an image popular? In *WWW*, pages 867–876, 2014.
- [Kulkarni *et al.*, 2013] Girish Kulkarni, Visruth Premraj, Vicente Ordonez, Sagnik Dhar, Siming Li, Yejin Choi, and Alexander C. Berg. Babytalk: Understanding and generating simple image descriptions. *IEEE TPAMI*, 35(12):2891–2903, 2013.
- [Landia *et al.*, 2012] Nikolas Landia, Sarabjot Singh Anand, Andreas Hotho, Robert Jschke, Stephan Doerfel, and Folke Mitzlaff. Extending folkrank with content data. In *RSWeb*, pages 1–8, 2012.
- [Li *et al.*, 2015a] Teng Li, Bingbing Ni, Mengdi Xu, Meng Wang, Qingwei Gao, and Shuicheng Yan. Data-driven affective filtering for images and videos. *IEEE TCYB*, 45(10):2336–2349, 2015.
- [Li *et al.*, 2015b] Xirong Li, Tiberio Uricchio, Lamberto Ballan, Marco Bertini, Cees G. M. Snoek, and Alberto Del Bimbo. Socializing the semantic gap: A comparative survey on image tag assignment, refinement and retrieval. *CoRR*, abs/1503.08248, 2015.
- [Lops *et al.*, 2013] Pasquale Lops, Marco de Gemmis, Giovanni Semeraro, Cataldo Musto, and Fedelucio Narducci. Content-based and collaborative techniques for tag recommendation: An empirical evaluation. *J. Intell. Inf. Syst.*, 40(1):41–61, 2013.
- [Lu *et al.*, 2014] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James Z. Wang. Rapid: Rating pictorial aesthetics using deep learning. In *ACMMM*, pages 457–466, 2014.
- [McParlane *et al.*, 2014] Philip J. McParlane, Yashar Moshfeghi, and Joemon M. Jose. "nobody comes here anymore, it's too crowded"; predicting image popularity on flickr. In *ICMR*, pages 385:385–385:391, 2014.
- [Murray *et al.*, 2012] Naila Murray, Luca Marchesotti, and Florent Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *CVPR*, pages 2408–2415, June 2012.
- [Page *et al.*, 1999] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. *Stanford Digital Libraries Working Paper*, 1999.
- [Pang and Lee, 2008] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, 2008.
- [Pang *et al.*, 2002] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *EMNLP*, pages 79–86, 2002.
- [Park and Zhang, 2015] Tae-Suh Park and Byoung-Tak Zhang. Consensus analysis and modeling of visual aesthetic perception. *IEEE TAC*, 2015.
- [Pedro and Siersdorfer, 2009] Jose San Pedro and Stefan Siersdorfer. Ranking and classifying attractiveness of photos in folksonomies. In *WWW*, pages 771–780, 2009.
- [Peng *et al.*, 2015] Kuan-Chuan Peng, Tsuhan Chen, Amir Sadovnik, and Andrew Gallagher. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *CVPR*, 2015.
- [Si *et al.*, 2009] Xiance Si, Zhiyuan Liu, Peng Li, Qixia Jiang, and Maosong Sun. Content-based and graph-based tag suggestion. In *ECML PKDD Discovery Challenge*, pages 243–260, 2009.
- [Sigurbjörnsson and van Zwol, 2008] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW*, pages 327–336, 2008.
- [Su and Khoshgoftaar, 2009] Xiaoyuan Su and Taghi M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence archive*, 2009.
- [Thomee *et al.*, 2015] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. The new data and new challenges in multimedia research, 2015. arXiv preprint arXiv:1503.01817.
- [Totti *et al.*, 2014] Luam Catao Totti, Felipe Almeida Costa, Sandra Avila, Eduardo Valle, Jr. Wagner Meira, and Virgilio Almeida. The impact of visual attributes on online image diffusion. In *WebSci*, pages 42–51, 2014.
- [van Zwol *et al.*, 2010] Roelof van Zwol, Adam Rae, and Lluís Garcia Pueyo. Prediction of favourite photos using social, visual, and textual signals. In *ACMMM*, pages 1015–1018, 2010.
- [Wu *et al.*, 2016] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, and Tao Mei. Time matters: Multi-scale temporalization of social media popularity. In *ACMMM*, pages 1336–1344, 2016.
- [Yamaguchi *et al.*, 2014] Kota Yamaguchi, Tamara L. Berg, and Luis E. Ortiz. Chic or social: Visual popularity analysis in online fashion networks. In *ACMMM*, 2014.
- [Yamasaki *et al.*, 2014] Toshihiko Yamasaki, Shumpei Sano, and Kiyoharu Aizawa. Social popularity score: Predicting numbers of views, comments, and favorites of social photos using only annotations. In *WISMM*, pages 3–8, 2014.
- [Zhang *et al.*, 2009] Yuan Zhang, Ning Zhang, and Jie Tang. A collaborative filtering tag recommendation system based on graph. In *ECML PKDD Discovery Challenge*, pages 297–306, 2009.
- [Zhang *et al.*, 2014] Luming Zhang, Yue Gao, Chao Zhang, Hanwang Zhang, Qi Tian, and Roger Zimmermann. Perception-guided multimodal feature fusion for photo aesthetics assessment. In *ACMMM*, pages 237–246, 2014.
- [Zhao *et al.*, 2016] Sicheng Zhao, Hongxun Yao, Yue Gao, Rongrong Ji, Wenlong Xie, Xiaolei Jiang, and Tat-Seng Chua. Predicting personalized emotion perceptions of social images. In *ACMMM*, pages 1385–1394, 2016.