Recommendation of Multimedia Items by Link Analysis and Collaborative Filtering

Davin Wong[1], Ella Bingham[2], Saara Hyvönen[3]

[1]Nokia Research Center, [2]HIIT University of Helsinki, [3]Fonecta Ltd.
davin.wong@nokia.com, ella@iki.fi, saara.hyvonen@fonecta.com

Abstract

We investigate two recommendation approaches suitable for online multimedia sharing services. Our first approach, UserRank, recommends items by global interestingness irrespective of user preferences and is based on the analysis of ownership and evaluation link structure. We also present a personalized interestingness algorithm that combines UserRank with collaborative filtering which enables a single parameter to control the degree of personalization in the recommendations. Our initial results from an informal user study are encouraging.

Introduction

The prevalence of broadband has made multimedia sharing a reality. Many online services have spawned over the recent years and sites like Flickr and YouTube are becoming favorite destination for ordinary people to publish, share and evaluate each other's photos and videos. As the collection of items grows exponentially, to expedite the retrieval of multimedia items, tags are frequently used to annotate photos and videos so that a simple keyword search can be used to return relevant items to a user. Unfortunately, tagging can be manipulated by spammers or malicious users by labeling their illegitimate items with irrelevant tags to gain exposure and publicity. In this paper, we address this issue by introducing a global interestingness measure for ranking items. We propose a link analysis algorithm called UserRank which analyzes the ownership and evaluation link structure available in a multimedia sharing site to produce global interestingness ranking for items.

Although tags are useful in helping users to locate relevant items, there are times when a user may simply want to explore the collection of items without any specific tags or topics in mind. While we could use UserRank to recommend items that are most globally interesting to the user, we can also try to predict what items the individual user may like based on her own portfolio of uploaded items or items she has evaluated in the past. To this end, we introduce a personalized interestingness measure for items. We use collaborative filtering to identify items belonging to users who share similar tag usage as the recipient of the recommendation. Then, we apply UserRank to score the items so that an item is recommended if its owner is similar to the recipient and the item itself is globally interesting.

UserRank

Our first objective is to assign global interestingness scores to all items in the collection. We surmise that users who own interesting items know what is interesting and their knowledge of interestingness influences them to pick out interesting items to evaluate. Hence, interesting items are bridged by interesting users (users who know what is interesting) through ownership and evaluation, and vice versa. We capture this relationship through the UserRank equation:

$u = B'Eu$ Interesting users own items evaluated by interesting users. (1)

where u is a column vector containing the interestingness scores for users; B is the ownership matrix whose element (i,j) equals 1 if row item i belongs to column user j, otherwise 0; E is the evaluation matrix whose element (i,j) equals $1/N_j$ if column user j evaluated row item i and N_j is the number of items evaluated by user j, otherwise 0. By evaluation, we make no distinction between actions taken toward one's own items or items of others.

Rather than ranking items directly, UserRank (1) ranks users first and from the ranked users assigns interestingness scores to items through their association with users. The (i, j) entry of B'E is nonzero only if any of the items owned by user i were evaluated by user j and the actual value is the proportion of user j's evaluations directed to user i. We can imagine the matrix B'E as depicting a competition for attention between the users.

To solve UserRank (1), we note that it resembles PageRank (Brin and Page, 1998) in form and so the computational steps are identical. We first replace any zero columns of the matrix B'E with a uniform probability vector so that it becomes column stochastic. Next, we apply the "teleportation" idea from PageRank to guarantee a unique dominant eigenvector u and the resulting system can be solved by the Power Method.

Once we obtained u, the ranking for users, we can rank the global interestingness of items by their association with users by:

$p_e = \hat{E}u$ Interesting items evaluated by interesting users.
$p_b = Bu$ Interesting items belong to interesting users.

where p_e and p_b are column vectors containing the global interestingness scores for items obtained through evaluation and ownership, respectively. The difference in the amount of evaluations accumulated over time between users can be quite high. For the evaluation matrix $\hat{E}$, instead of normalizing the elements in a column by the total number of evaluations by a user, we can normalize by N^k where N is the number of items evaluated by the user. We found that the range $0 \leq k < 0.5$ gives acceptable results. To obtain the final global interestingness ranking for an item, we combine p_e and p_b with a convex sum:

$$\text{rank} = \gamma p_b / |p_b|_1 + (1-\gamma) p_e / |p_e|_1 \quad (2)$$

where γ is a real number between 0 to 1. We can either use a cutoff threshold to select items to recommend or pick the top M items. We can vary γ to control the significance of prominent ownership. When $\gamma=1$, the recommended items are sorted by popularity of the owners in descending order. When γ is small, the recommendation tends to be less dominated by any single user. By choosing $\gamma>0$, a newly submitted item that has not been evaluated by others can still receive positive ranking and has a chance of being recommended.

Link-Based Collaborative Filtering

Our second objective involves making personalized recommendation to individual users based on their preferences. We use collaborative filtering (Adomavicius and Tuzhilin, 2005) to recommend the items of users who share similar interests as the recipient. For our problem domain, the user-to-item ratio can be quite low and chances of any two users having rated many of the same items are miniscule so we cannot rely on common items evaluated by the users to compute user similarity. To avoid this so-called sparsity problem, we compute similarity by applying the vector space model on the tag frequencies of items owned by users. Our assumption is that the set of tags chosen by a user to describe her items bears the telltale sign of her preference. Another possibility is to use the tags associated with items evaluated by users to compute similarity. While any two users are similar if their tag usage overlaps, two users without overlapping tag usage may also be indirectly similar if there exists a third user with overlapping tag usage to both users. This third user acts as a bridge in connecting the two otherwise dissimilar users.

In matrix notation, we first define the tag-user matrix A such that each column in the matrix lists the tag frequencies of the items owned by a user and the columns are normalized to Euclidean norm 1. Then we compute the aggregated user similarity matrix D by summing over the indirect user similarities:

$$D = \mu^{-1}(A'A) + \mu^{-2}(A'A)^2 + \ldots + \mu^{-n}(A'A)^n \quad (3)$$

where μ is a constant for attenuating the significance of indirect similarity over long chains of intermediate users. Each $\mu^{-k}(A'A)^k$ term holds the attenuated similarity between users at k hops away. We choose some small n and set μ to greater than the largest eigenvalue of A'A. The dampening factor μ allows us to control the neighborhood size of similar users and is used to control the degree of personalization. When μ is equal to the largest eigenvalue of A'A, all users will be similar to each other as long as there is a path of length less than n connecting them. When μ is much greater than the largest eigenvalue of A'A, similarity is dominated by direct overlapping tag usages between users. To recommend items to a user, we define the ranked ownership matrix P whose element (i,j) is 0 unless column item j belongs to row user i in which case the value is set to rank_j computed from Equation 2. Each row of the matrix product DP then holds the items ranking specific to a user based on the similarity between the user and item owner as well as the interestingness of the item.

Experimental Results

We conducted an informal user study based on a small data set crawled from Flickr with 2524 users who collectively own 2,177,103 photos. We gathered evaluation and ownership information to compute UserRank, and also tag usage data to compute user similarity. Of the photos collected, about 35% are commented by users from the set. On average, each user commented 302 photos and marked 120 photos as their favorites. We asked 19 volunteers from the data set to rate the top 100 items recommended by the algorithm and the results seem to suggest that both global and personalized interestingness measures were able to recommend photos of certain aesthetic quality and general interestingness to the evaluators. In terms of personal appeal of the recommendations, our test group was too small to draw any definitive conclusion, therefore, we plan to conduct a more thorough evaluation of the proposed methods in the future and also compare the algorithm to others such as EigenRumor (Fujimura et al, 2005).

References

Adomavicius G., and Tuzhilin A. “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”. IEEE Tr Knowl Data Eng, 17(6), pp. 734-749, 2005.

Brin S., and Page L. “The Anatomy of a Large-Scale Hypertextual Web Search Engine”. Computer Networks and ISDN Systems, 1998.

Fujimura K., Inoue T., and Sugisaki M. “The EigenRumor Algorithm for Ranking Blogs”. Workshop on the Weblogging Ecosystem, May 2005.