WEB VIDEO RECOMMENDATION AND LONG TAIL DISCOVERING

Xiao Wu[1, 2], *Yongdong Zhang*[1], *Junbo Guo*[1] *and Jintao Li*[1]

[1]Key Laboratory of Intelligent Information Processing,
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[2]Graduate University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

Given countless web videos available online, one problem is how to help users find videos to their taste in an efficient way. In this paper, to facilitate user's browsing we propose relevant and exploratory recommendation algorithms utilizing multimodal similarity and contextual network to organize web videos of various topics. Comparison experiments demonstrate proposed approach generates more accurate video relevancy. And our method is more flexible in discovering user latent interests in long tail videos.

Index Terms— Recommendation, contextual network, long tail, multi-criterion ranking, visual duplicate detection

1. INTRODUCTION

Online video sharing has achieved great success recent years. Via helping customers enjoy web videos, video sharing websites (e.g. YouTube) attract much attention from both users and advertisers. However, how to help millions of users find their favorites from billions of videos is a very challenging issue. This issue is essentially a problem of information overload [1]. Recommendation system provides a solution for this problem. Given an object (e.g. a video in Fig.1) browsed by the user, recommendation algorithm generates the recommend list of objects (e.g. similar videos) for the user to choose. Under circumstance of information overload, people usually prefer being recommended to searching. On anther hand, recommendation plays key role in long tail mining [2], which include clustering objects into categories, comparing various products and catering various non-mainstream demands [2].

Recommender systems, which have been widely used in e-business, Netnews and online music sharing [3], aim to lead user to find his desired products. Three categories of algorithms have been developed for recommender systems [1], i.e. the widely used global ranking (GR), content based method (CB) and collaborative filtering (CF). GR methods rank objects using single criteria such as video viewed time. CB methods analyze the correlation between objects. CF methods study user behavior based on abundant user transaction histories. For web video sharing, since users usually browse video websites in a casual manner without login, the transaction histories are incomplete and hence CF methods are disabled. Though widely used in current video sharing websites, GR based on single criterion and CB based on title or tag similarities do not perform well in video recommendation, e.g. recommend merely videos of mainstream or return unrelated videos of similar title.

While most academic researches focus on video search, few works study how to recommend videos to user. In this work, we address the problem of web video recommendation and convert it into tasks of relevant recommendation and exploratory recommendation. Our approach based on the contextual network of web videos aim to discover semantic "cliques" (i.e. videos of one latent topic) using spectral partitioning. Relevant recommend list is generated using videos within "clique" and long tail videos are discovered in neighboring "cliques". Comparison experiments on real world video dataset collected from YouTube demonstrate the advantages of our approach to current used algorithms.

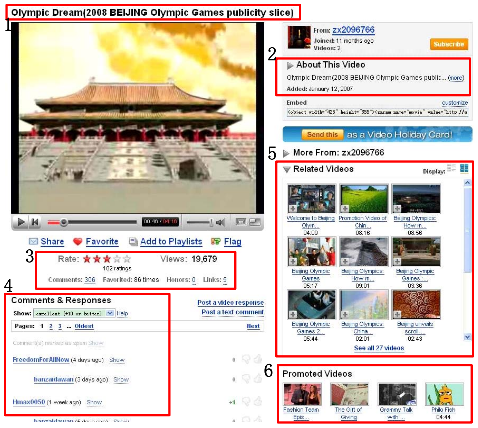
Olympic Dream(2008 BEIJING Olympic Games publicity slice)
1
From: zx2096766
Joined: 11 months ago
Videos: 2
Subscribe
2
About This Video
Olympic Dream(2008 BEIJING Olympic Games public... (more)
Added: January 12, 2007
Embed
customize
<object width="425" height="355"><param name="movie" value="http://w
Send this as a Video Holiday Card!
5
More From: zx2096766
Related Videos
Display:
Welcome to Beijing Olym... 04:09
Promotion Video of Chin... 08:16
Beijing Olympics: How m... 08:56
Beijing Olympic Games 05:17
Beijing Olympics: How m... 09:01
Beijing Olympic Games :... 03:36
Beijing Olympic Games 2... 05:44
Beijing Olympics: China... 02:01
Beijing unveils scroll-... 02:43
See all 27 videos
00:46 / 04:16
Share
Favorite
Add to Playlists
Flag
3
Rate:
Views: 19,679
102 ratings
Comments: 306 Favorited: 86 times Honors: 0 Links: 5
4
Comments & Responses
Post a video response
Post a text comment
Show: excellent (+10 or better) Help
Pages: 1 2 3 ... Oldest
Next
Comment(s) marked as spam Show
FreedomForAllNow (4 days ago) Show
banzaidawan (3 days ago) Show
Hmax0050 (1 week ago) Show
6
Promoted Videos
Fashion Team Epis...
The Gift of Giving
Grammy Talk with ...
Philo Fish 04:44

Fig.1. A sample web video page from YouTube.com
The page consists of 5 contextual layers, denoted by the red boxes.
(1) video title (2) description and tags (3) user blog links and rating
(4) user comments (5) related recommendation (6) promoted video list

2. WEB VIDEO CONTEXTUAL NETWORK

The motivation of building the contextual network is utilizing content and context information to calculate the correlation between web videos of a sharing website.

2.1. Incremental construction of the network

The first objective is utilizing explicit links on video pages (as in 5th layer of Fig.1) to initialize the structure of network. Similar network or graph is also used in several previous works [5] [6]. The difference is there is no explicit link could be used in [5] [6] hence pairwise similarity has to be calculated which is costly for large scale data. Our idea is more similar to [7], both utilize explicit links on web pages.

Since it is impossible and unnecessary to collect all the millions of web videos in the website, we propagate the network in an incremental manner. Given a randomly selected web video as the seed, the network propagates abided to the breadth priority norm, i.e. given a video page, it first include all the explicit linked videos as new nodes.

Formal definition of the contextual network is given here, **Definition**: Given a set of web videos, a contextual network is defined as an directed weighted graph denoted as N = {V, E}, which consists of node set V and edge set E. V = {n_1, n_2, …, n_i, …, n_m}, node n_i denotes ith video. E = {e_1, e_2, …, e_j, …, e_n}, edge e_j denotes correlation between two videos.

2.2. Condensing network using multimodality similarity

Noted that there are a few videos with hundreds of explicit links, N should be condensed to eliminate noises. The content based similarity is calculated as follow,

$$SIM(n_i, n_j) = \alpha * TS + (1-\alpha) * VS \tag{1}$$

A free tunable parameter α is introduced for linear combination of textual and visual similarity, which is denoted as TS and VS separately. First textual annotations on video page (in first and second layer in Fig.1) are stemmed and segmented into scatter words. TS is calculated using cue work clusters and cosine similarity discussed in [6]. Second, near duplicate detection scheme [8] is adopted to rerank videos using visual similarity. And a threshold is decided to filter out noise links of lower similarity. Though utilize the initial link structure of video website, we refine the network structure to finally improve accuracy. As discussed above, comparison of various products (videos) is one of the essential features of long tail strategy.

2.3. Incorporating contextual information and converting network to undirected graph

Contextual information, e.g. user opinions, viewed times and video age (as in third layer of Fig.1), which enables us to determine the mainstream degree of web videos, are incorporate to calculate the weight of node n_i, WN(n_i).

To apply spectral techniques, N is converted to undirected weighted graph (denoted as G). During the conversion content and context information are incorporated together to compute the weight of undirected edge, denoted as WE(e_j). WN(n_i) and WE(e_j) are calculated as follow,

$$WN(n_i) = \beta_1 * \log(viewed\_times) + \beta_2 * \#comments + \beta_3 * novelty$$

$$\text{s.t. } \sum_{i=1}^{3} \beta_i = 1 \tag{2}$$

$$WE(e_j) = \gamma * \sum_{k=i,j} WN(n_k) / \#Link_k + (1-\gamma) * SIM(n_i, n_j) \tag{3}$$

where #comments in (2) denotes number of user comments. Novelty is designed to evaluate the aging factor of video, calculated as 1/|current_date – video_release_date|.

The key step of the conversion is illustrated in (3), where WE(e_j) is computed taking both content and context correlation into considerations. #Link denotes the number of forward links on video page (in 5th layer in Fig.1). Similar to [7], WN(n_i) has contribution to all forward edge weights. This is essentially a heat diffusion process. “Hot” nodes (videos) are surrounding by solid edges and the coupling degree of related videos enhances. All the variables in these formulas are normalized to eliminate the scale difference.

A simplified example of G is shown in Fig.2 (a). The parameters we use in this work is tuned and fixed, i.e. α =0.5, $\beta_1 = \beta_2 = \beta_3$ =0.3. Especially, γ is set to 0.3 to slightly reduce the influence of mainstream.

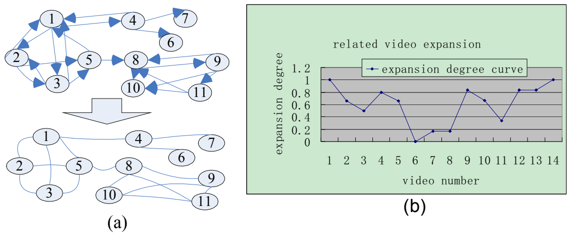

Fig.2. (a) converting N to G, related videos are distributed in several “cliques” (b) Dex curve of N, consecutive minima denote the existence of “clique”

3. SPECTRAL TECHNIQUES ON UNDIRECTED WEIGHTED GRAPH

Given the undirected graph G, we aim to discover a series of “cliques” where videos are relative focus on the same topic. By partitioning G into G_i, videos belong to related semantic “cliques” could be clustered. Clustering objects (e.g. videos) into various categories is another important key of long tail strategy as discussed in the introduction.

3.1. Spectral graph partitioning and clustering

Discovering videos “clique” could be treated as a problem of spectral graph partitioning from the view of global graph, or a problem of spectral clustering from the view of scattered data (videos). Given G and symmetry weight matrix W (W_{ij} presents the correlation between two videos), a diagonal weight matrix D is computed, where $D_{ii} = \sum_j W_{ij}$.

D_{ij} denotes the sum of weights of node n_i. To partition G into semantic "cliques" we adopt the normalized cut algorithm [4], which could be reduced to a generalized eigenvector problem as follow,

$$Ly = \lambda * Dy \quad (4)$$

Where L=D-W, denoted as the Laplacian matrix. We can easily get the following equivalent formula,

$$D^{-1}Wy = (1-\lambda)y \quad (5)$$

Denote $D^{-1}W$ as T, i.e. the transition matrix. T_{ij} could be explained as the probability of user jumping from web video i to j. The partitioning problem is solved by calculating y, the right eigenvector of T. Similar to [4], K eigenvectors corresponding to K minimal eigenvalue are preserved and Kmeans algorithm is conducted to determine clustering.

3.2. Determining the K for graph partitioning

One of the most important issues of spectral clustering is how to determine the number of clusters K, which is basically depend on the particular problem. This is also a key parameter for many other clustering algorithms in literature, such as Kmeans. It can be observed in Fig.2 (a) that completely connected local structure has local maximal weight density, which verifies our resumption that there exist many video "cliques" to be discovered. Based on this observation, we propose dynamic expansion degree (Dex) to evaluate the propagation speed of topic in the incremental process, which could be utilized to guide us how to determine K. The Dex of node n_i in N is computed as below,

$$Dex(n_i) = \# N_{new} / (1 + \# N_{pre}) \quad (6)$$

Where $\#N_{new}$ denotes number of new propagated nodes. $\#N_{pre}$ denotes number of edges linking n_i to previous generated node n_j or to neighboring nodes of n_j. As shown in Fig.2 (b), the curve of Dex contains several local minima (trough) which present the large back-link degree (i.e. videos in "clique" link each others) and indicate the existence of a "clique" of G. The number of consecutive local minima is used as K. On another hand, the peaks of Dex curve present fast transition from one topic to another.

4. STRATEGIES FOR RELEVANT AND EXPLORATORY RECOMMENDATION

In real applications, given a web video watched by the user, current video sharing websites conduct relevant recommendation and global recommendation (as shown in Fig.1) for user to browse more videos. Related videos are searched merely by text based similarity. Global recommended videos are ranked and selected using single criterion such as view times which indicate the mainstream degree of videos.

In contrast, we divide the task of video recommendation into relevant and exploratory recommendation. Three rules are proposed for recommendation algorithm design:

- Related video recommendation should take both textual and visual similarity into consideration to eliminate false positive textual connections.
- Exploratory recommendation should function as the guidance to help user discover latent interests in non-mainstream videos (long tail).
- Local ranking (LR) scheme is adopted for relevant and exploratory recommendation, i.e. videos in lists are reranked based on mainstream popularity (i.e. $WN(n_i)$).

Based on video "clique" G_i partitioned from G, two strategies are designed subjected to the 3 rules. Given query video Vq in G_i, the relevant recommendation list is constructed by return neighboring nodes of top LEN weighted edge. LEN denotes the length of list. For exploratory recommendation, each "clique" is represented using the node of largest $WN(n_i)$ to construct a new graph denoted as GC. Edge weight of GC is computed by summing up all the edge weights between two neighboring "cliques". Then exploratory recommendation list is generated in the same way as relevant recommendation. It conducts topic transition to help user browse farther. Videos in the lists are reranked according to $WN(n_i)$ to conduct LR.

5. EXPERIMENTS AND DISCUSSIONS

5.1. Experiment setup

Test videos are extracted from YouTube.com. As discussed above, we adopt an incremental manner to build contextual network N. How to decide the size of N is another important issue for test data setup. In our work, we construct N at the scale of 300 videos. Natural logarithms of video view times are approximately subject to Gaussian distribution, which indicate that the data set is selected without loss of generality. Tunable parameters for N construction are tested and the parameter combination of best performance discussed in Section 2.3 is adopted. To evaluate the performance, 10 videos in N are randomly chosen as queries under constrain that each video query has more than 30 neighboring nodes to facilitate the experiment.

For the experiment of relevant recommendation, two undergraduate students as the assessors are required to watch the query and recommended videos to generate grand truth (GT) by labeling results as related or unrelated. We use mean average precision (MAP) calculated by comparing result list R with GT of the queries. The average precision (AP) for each list of a query is calculated as follow,

$$average\_precision = \frac{1}{N_r}\sum_{i=1}^{N_r}\frac{i}{Related\_Potision(i)} \quad (7)$$

Where N_r denotes number of related video in result list. Related_Position(i) denotes the position of i^{th} related video.

For the experiment of exploratory recommendation, we propose a simple measure called "long tail strength" (LTS) to compute the proportion of non-mainstream in the result list, which is calculated as follow,

$$LTS = | R \cap NM | / | R | \qquad (8)$$

where NM denotes the non-mainstream video in recommend list, i.e. videos not in the top viewed list of YouTube. LTS represents the ability of long tail discovering of recommender system. The larger LTS is, the more long tail videos are discovered in the list.

5.2. Performance of relevant recommendation

To conduct the comparison experiments, a baseline run based on textual similarity (i.e. α =1) graph is build and used. In this run we construct a KNN graph using the explicit links on video page and K is set as 30 (i.e. 30 neighboring nodes for each node at most). The KNN graph is constructed in the same manner as N. In addition the real world performance is introduced to compare as another run. Therefore, given a video query we obtain 3 relevant recommendation lists. By comparing lists of 10 queries to GT, the MAP is computed at different length of the list.

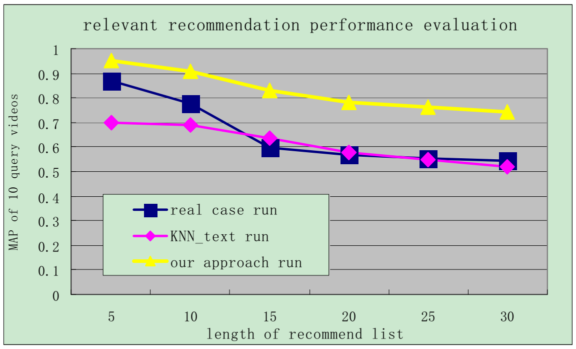

Fig.3. MAP curves of three runs for relevant video recommendation

Fig.3 demonstrates three MAP curves for the 3 runs, which shows that when length of results list is larger than 15, distinct robustness of our approach reveals. The video results of our run are relatively focused on identical topic. The main reason lies on the elimination of false positive connections between videos. First, during the construction of N we introduce visual duplicate detection to enhance the correlation between near identical videos, which widely appear [9]. Second, video clustering based on spectral partitioned "cliques" narrows the range of related video searching hence reduces the chance of false matching. In additional, related video selection within local structure reduces the computational complexity greatly.

5.3. Performance of exploratory recommendation

Different from traditional text information retrieval which avoid the problem of topic transition, exploratory recommendation performs active and rational topic transition to lead user to discover latent interests. We compare our exploratory recommendation lists of 10 queries to "Promoted videos" lists in YouTube. Length of the list is fixed to 4 as it is in YouTube. The average LTS is 0.54 of our approach and 0.32 of YouTube. "Promoted videos" in YouTube are selected from new and popular videos while our long tail videos are selected from topics ("cliques") neighboring to query. As an effective solution for information overload, our strategy aim to narrow the range of user choice according to which video he watches and lead user to find videos to their taste efficiently.

6. CONCLUSION AND FUTURE WORKS

Recommendation strategy for web video sharing is of abundant practical application. We address the problem of how to help user to copy with information overload and propose relevant and exploratory recommendation strategies based on multi-criterion and contextual similarity. Future research will focus on adopting new graph algorithms on large scale data using the YouTube results as the baseline.

7. ACKNOWLEDGEMENTS

This work was supported by the National Basic Research Program of China (973 Program, 2007CB311100), National High Technology and Research Development Program of China (863 Program, 2007AA01Z416), National Nature Science Foundation of China (60773056), and the Beijing New Star Project on Science & Technology (2007B071).

8. REFERENCES

[1] G. Adomavicius, A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Trans. on KDE 17(6), 734–749 (2005).

[2] Chris Anderson, *The Long tail*: BOOK-STUDIO Press, 2006.

[3] Amazon.com, eBay, Levis, Ski-europe.com, CDNOW, CoCoA, Ringo, Moviefinder.com, MovieLens, Reel.com, GroupLens.

[4] J. B. Shi and J. Malik, "Normalized cuts and image segmentation", IEEE Tran. PAMI 22(8), 888-905, 2000.

[5] X. He, W. Y. Ma, and H. J. Zhang, "ImageRank: spectral techniques for structural analysis of image database", ICME, 2003.

[6] W. H. Hsu, L. S. Kennedy, S. F. Chang, "Video Search Reranking through Random Walk over Document-Level Context Graph", ACM multimedia, 2007.

[7] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web", Stanford Digital Library Technologies Project, Technical report, 1998.

[8] Yantao Zheng, Shi-Yong Neo, Tat-Seng Chua, Qi Tian, "Fast Near-duplicate Keyframe Detection in Large-scale Corpus for Video Search", In IWAIT 2007, Bangkok, 8-9 Jan 2007.

[9] X. Wu, A. G. Hauptmann, C. W. Ngo, "Practical elimination of near-duplicates from web video search", ACM Multimedia, 2007.