External Links of Video Sharing using RFP Recommendations

Sandhya Shinde
PCCOE-IT Dept.

Prasad Kajale
PCCOE-IT Dept

Shivprasad Pawar
PCCOE-IT Dept

ABSTRACT
Now a day's popularity of video sharing site is increased. People can watch videos from sites and also interested in relevant links videos which are suggested by social video sharing sites. To increase popularity of video external links concept is used. Now in video sharing sites through external links video or audio contents can be embedded into external web sites. User can copy the URL(uniform resource locater) of that embedded link and post on their own blog or website. In this paper intention is study of relevancy of videos and & increase the popularity and measure the quantification. With the results collected from two major video sharing sites like YouTube & Youku.Then observed that these links have a various impact on popularity. Overall, videos which are collected from external links are analyzed also accuracy & popularity is measured.

Keyword
Video sharing, YouTube, Youku, External links, frequent item set; data mining.

1. INTRODUCTION
Media organizations are the mediator for distributing media contents through the regional markets .For example, Video distribution. Every user do not think the same way, they could not access the same contents in entire world. User Generated Content sites are most popular now a days. In these sites user can actively upload their own content as well as can access their information. Facebook[9], twitter, Flicker, Video sharing [8]sites like YouTube are the examples of UGC[7] sites. Basically, this paper focuses on video sharing sites like YouTube[1] and Youku[2]. The reason behind popularity of UGC (User generated content) sites is information distributed much faster through sites. If consider YouTube then it provides various functionality which includes related video links which arrange videos by related topic. To increase the popularity of video distribution many sites introduce External Links on other websites. Consider You Tube, some embedded links are provided there for each video. Then, question is how popularity of videos is increased? User can copy one of these embedded link's URL and post it into their personal web pages or blogs. When people watch video through external links then count goes through You Tube and popularity of site increases. Clearly, external links helps to embed videos in non-video sharing sites to attract views.

2. MOTIVATION
Youku & You Tube are two UGC (User generated content) sites. These sites focus on user-to-user, user-to-video or video-to-video relationship. Users can get embedded link URL & post this into their website, forums or blogs. And also collected number of count in the user account. System used these two tables for calculating the recall and precision in this dataset. High recall means this algorithm returned most of the relevant results

3. LITERATURE SURVEY
The relationship between popularity and locality of online YouTube videos. It illustrate whether YouTube videos publically display geographic locality of interest, with views arising from a spatial area rather than from a global one. This analysis is done on more than 20 millions YouTube videos, uploaded in one year from different regions. It find that about 55% of the videos have more than 75% of their views from a single region. By relating locality to viralness show that social sharing generally become wider geographic reach of a video. it gets stuck in a geographic region. Finally analyze how the geographic properties of a videos views evolve on a daily basis during its life time, providing new insights on how the geographic reach of a video change as its popularity peaks and then fades away.

A video sharing service allow user-generated video clips to be uploaded and user of the service to view, rate, and comment on uploaded video. One common aspect of these services is the ease of uploading, searching, and viewing videos. Generally, these sites allow content producers to upload content encoded using any commonly used codec. That uploaded video is converted by the service provider to a common format , thus enabling most Web users to view the itemset without search for different codecs to view different videos. Availability of a many number of diverse videos, the ease of viewing at the click of a button, and in the ability to form social group with content uploaders and viewers alike also contributes towards the increased popularity of these services.(advantage and disadvantage some algorithm Described in table 1)

Table 1: Literature review table

Paper	Advantages	Disadvantages
Vivisecting YouTube :An active measurement study	provides a method to count the total number of YouTube videos by leveraging the video ids	videos would be biased towards the popular videos,
I tube, YouTube, everybody tubes: Analyzing the largest user generated content video system	video sharing[8] sites widely spans to user-to-user, user-to-video, and video-to-video relationship 2>alters the skewness of popularity, or	1>the impact of content aliasing and illegal uploads, which could hamper the future success of UGC services

	breaks the power-law behavior for very popular contents	
YouTube around the world: Geographic popularity of videos	1>online video consumption appears geographic locality of interest 2>availability of finer-grained geographic information about video items and video views would make spatial probabilistic models possible	1>delivery networks to recommendation and discovery engines, 2>delivery networks to recommendation and discovery engines,
Track globally, deliver locally: Improving content delivery networks by tracking geographic social cascades	improve multimedia content delivery in YouTube is suggested	Popular videos are correlated with virus.
Characterizing web-based video sharing workload	video popularity distribution	Cache clean all details gone
Collaborative filtering for orkut communities :Discovery of user latent behavior	1>association rule mining (ARM), which discovers associations between sets of communities that are shared across many users	to take multi-order rules into consideration rather than just first order rules.
Hot Today, Gone Tomorrow: On the Migration of MySpace Users	Characterization of the OSNs' friendship structures inferred from single Snapshots taken at a particular point of time.	Not working on multiple snapshot

4. PROPOSED SYSTEM

Analyzing external links of videos using Association Rule Mining[4]. In a proposed method, finding & analyzing on number of hits videos from external links with different categories, personalization of user. Also find relevant video links with relative analysis of Apriori algorithm and parallel RFP(relational frequent pattern)-Growth algorithms.(proposed system architecture described in fig 1)

In this system how module to be process

- User login and register
- Admin login and register
- Shared link one or more user Comment on blog
- User likes video and shared.
- Showing result categorywise and sub catagorywise.
- Identification of external links.

Association rule mining[3] is popular and important technique of data mining. It is intended to discover the interesting patterns from the large transaction itemset, that is, frequent patterns and association rules.

3.1 Association Rule Mining:

Association rule mining[3] is one of the unique techniques of data mining. it is the most common form of local pattern discovery in unsupervised learning system. It provide as useful tool for finding correlations between items in large databases item. The term used in these rule are

3.1.1 Support:

The support[6] supp(A) of an item set A is defined as the proportion of transactions in the data set which contain the itemset.

$$Supp(A) = \frac{No.oftrans.whichcontaintheitemset}{Totalno.oftrans.}$$

3.1.2 Confidence:

The confidence[6] for an association rule A implies B is the ratio of number of transaction that contains $A \cup B$ to number of transaction that contains X.

$$Conf(A \rightarrow B) = \frac{Supp(A \cup B)}{Supp(A)}$$

Large Item Set: A large item set is an itemset whose no. of occurrences is above a threshold or support.

3.2 RFP (Relational frequent pattern)

1) Construct a FP[5] Tree-Build a compact data structure called the FP-tree. It built using 2 passes over the data-set.

Pass 1:

- Find support

- Remove infrequent items.

- Sorting.

Pass 2:

- Reads one transaction at a time.

- Fixed order is used

- Pointers are maintained

2) Remove frequent item set from tree directly.

Since FP growth algorithm is advanced algorithm of Apriori Because if the database sizes is improved Apriori is not Efficient. So using improved FP growth is called RFP growth to avoid generating intra-property frequent item sets, and to further boost its efficiency, implement its Map-Reduce[10] version with extended prune strategy.

3.3 Mathematical Model

Let system ,

S={U,V database ,CV,I,ELV,ILV,RFP-Growth, A,U V Details}

U is no. of users ={U1, U2, U3....}

V Database=No. of videoes in database ={VD1, VD2, VD3....}

V Details={Upload data, no of view, Related IL,ELV,CV}
Upload date= find age of video

ELV={External links of videoes}

ILV={Internal links of video}

No of views NV={NV1, NV2....}

CV is category of video={CV1, CV2, CV3....}

RFP Growth={CV,N,U,T,Q,H,N}

HT=Header table{HT1, HT2}

QL=node links={QL1, QL2,....}

T=Tree generated by link nodes N =Nodes{N1, N2,....}

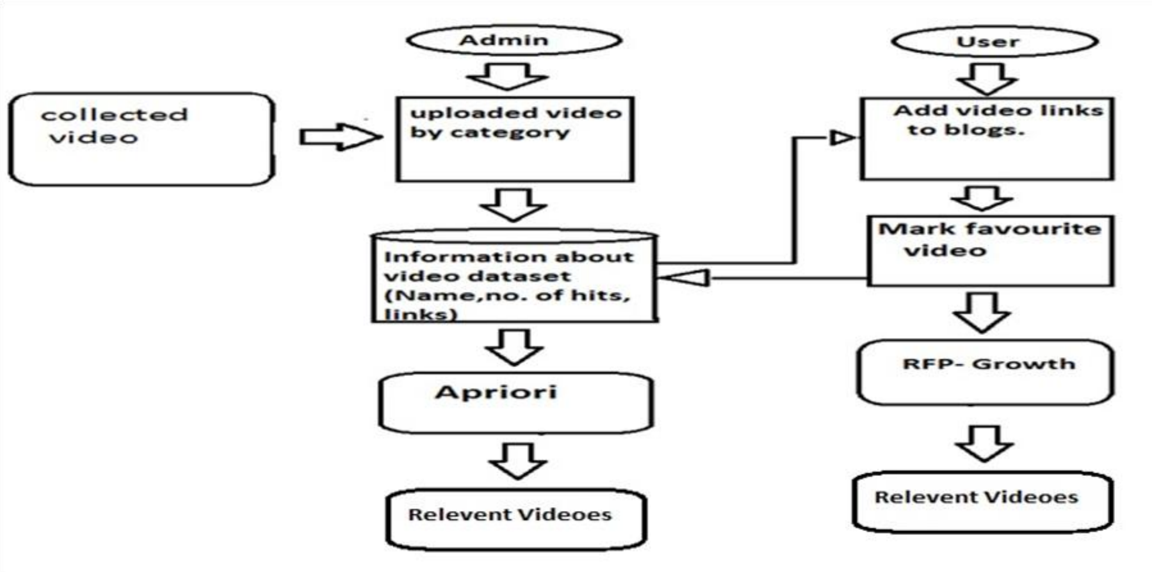

Fig 1. System Architecture

5. EXPERIMENTAL SETUP

Required Operating System: Windows 7, IDE: Net Beans 8.0, Programming Language: Java, Any Processor above 500 MHz, RAM 1 GB, Hard Disk 10 GB .

5.1 Dataset

In proposed method calculate results. And create database of videos of following categories just like sport ,funny, song ,documentary, Technical and etc. and respectively .To use database of no. of videos overall 100 approximately.

5.2 Result

Used compact data structures and eliminate repeated database scan. Using collaborative filtering[6] will used. Finding every occurred frequent itemset satisfying minimum support and minimum confidence. Hence this system Showing result categoriwise and sub-catagoriwise also showing personalized Identifiaction of external link and recommendation. to avoiding intraproperty and personalization of user account and share video link.

Recall and precision[12]values have been measured for dynamic programming and model based user item dataset and the number of items which are all presents in all data. And also collected number of count in the user account. So used these both tables for measuring the recall and precision in this dataset. High recall means that an algorithm returned most of the qualified result. High precision means that an algorithm returned more relevant results than unessential

$$Precision = \frac{|\{Relevantvideo\} \cap \{Retrievedvideo\}|}{|\{Retrievedvideo\}|}$$

$$\boldsymbol{Recall} = \frac{|\{\boldsymbol{Relevantvideo}\} \cap \{\boldsymbol{Retrievedvideo}\}|}{|\{\boldsymbol{Relevantvideo}\}|}$$

6. CONCLUSION

It is an important aspect of video sharing sites, the external links. The external links provide a unique way for the video sharing sites to accelerate the distribution of the videos and the external links can play a non trivial role both in terms of the number of external links on a video, and the of views contributed to the video. The external links have quite different effect on YouTube and Youku, the correlations of the external links and the internal related video links .From the experimental data items presented it can be concluded that the FP-growth algorithm behaves better than the apriori algorithm[11][12].

For Increase scalability, Naïve Bayes classifier can be used in future for categorization and personalization of videos.

7. ACKNOWLEDGMENTS

It gives us pleasure in presenting the preliminary project report on 'External Links of Video Sharing Using RFP and Recommendation. I would like to thank my internal guide Prof. Sandhya Shinde for giving me all the help and guidance I needed. I am grateful to them for their kind support. Their valuable suggestions were very helpful .I am also grateful to Dr.Sudhip Thepade, Head of Information Technology Department, PCCOE for his indispensable support, suggestions.

8. REFERENCES

[1] YouTube. [Online]. Available: http://www.youtube.com.

[2] Youku [Online]. Available: http://www.youku.com.The data sets of external links from YouTube and Youku.

[3] Q. Ruan, H. Lin and H. Tan. Research on the approach to optimize book exhibition based on mining association rules. International Journal of Advancements in Computing Technology, 2012. 4(16): pp. 500-507.

[4] H. He. Analysis ofAssociation Rules in Book Circulation. Library Journal, 2011. 7(30): pp. 63-68.

[5] Aggarwal, C.C., Li, Y., Wang, J. and Wang, J. (2009) Frequent pattern mining with uncertain data, International Conference on Knowledge Discovery and Data Mining, Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACMPp. 29-38, ISBN-978-3-642-13657-3.

[6] W. Chen, J. Chu, and J. Luan, "Collaborative filtering for orkut communities: Discovery of user latent behavior," in Proc. ACM WWW'09, Madrid, Spain, Apr. 20–24, 2009.

[7] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in Proc. ACM IMC'07, San Diego, CA, Oct. 24–26, 2007.

[8] S. Mitra, M. Agrawal, A. Yadav, N. Carlsson, D. Eager, and A. Mahanti, "Characterizing web-based video sharing workload," ACM Trans. Web, vol. 5, no. 2, May 2011.

[9] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in Proc. ACM WWW'10, Raleigh, NC, Apr. 26–30, 2010.

[10] X.Y. Yang, Z. Liu and Y. Fu. MapReduce as a programming model for association rules algorithm on Hadoop. 3rd International Conference on Information Sciences and Interaction Sciences (ICIS 2010). 2010. Chengdu, China.

[11] J. Cryans, S. Ratte and R. Champagne. Adaptation of apriori to MapReduce to build a warehouse of relations between named entities across the web. 2nd International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA 2010). 2010. Menuires, France.

[12] M. Lin, P. Lee and S. Hsueh. Apriori-based frequent itemset mining algorithms on MapReduce. 6th International Conference on Ubiquitous Information Management.