Comparative Analysis of External Links of video Using Apriori and RFP

Pooja D.Mahamuni
ME Computer Student
Department of Computer Engineering
JSCOE PUNE
Pune, India
pooja.mh4709@gmail.com

Prof. H. A. Hingoliwala
Assistant Professor
Department of Computer Engineering
JSCOE PUNE
Pune, India
ali_hyderi@yahoo.com

Abstract— Recently popularity of video sharing site is increased. People can watch video from any other site other than video sharing site. To increase popularity of video external links concept is introduced. Now in video sharing site through external link is video or audio content can be embedded into external website. User can copy that URL of that embedded link & post on their own blog or website. In this paper our intension is study of relevancy of videos and & increase the popularity & measure the quantification. With the results collected from two major video sharing site like Youtube & Youku we observed that these links have an various impact on popularity. Overall, videos which are collected from external links are analyzed and accuracy & popularity is measured

Keywords-*Video sharing,Youtube Youku,external link,UGC*

*****

I. INTRODUCTION

Historically, Media organization is the mediator for distributing media contents through the regional markets.Video distribution is one of the example of it. Every users does not have same mentality, they could not access the same content all over the world. User Generated Content (UGC) sites are most popular now a days. In these sites along with accessing the information user can actively upload their own content. Facebook, twitter, Flicker, Video sharing sites are the examples of UGC sites.

Among these sites this paper will emerging the concepts of video sharing sites which comparatively focus on results of You Tube & You ku. The reason behind popularity of UGC sites is information distributed much faster through this sites.

If we consider the You Tube then there are various functionality in that. One of them is related video links which arranges videos by similar topic. To increase the popularity of video distribution these sites introduce External Links. Consider Youtube, embedded links are provided there for each video. Then how popularity of videos are increased? User can copy one of these embedded link's URL and paste into their either on personal web pages or blogs. When people watch video through that external links then count goes through

You tube and thus popularity is increased. Clearly, external links are helpful to embed videos in non video sharing sites to attract views.

Motivation:

Youku & You Tube are two UGC sites.In this, focus is on user-to-user, user-to-video or video-to-video relationship among videos.Users can get embedded link URL & paste this into their website,forums or blogs.

II. LITERATURE SURVEY

There is relationship between locality & popularity of videos. Geographic locality of interest occurs in online video consumption. It uses new measures to quantify their popularity distribution across different regions.[1].

Analysis of popularity evolution content duplication & distribution of UGC video content. We understand the growth of UGC sites & its impact on behavior of user, infrastructurer[2].

We used Association Rule Mining technique to discover the correlation between set of items. From supermarket analysis that which two items are purchased together is motivation.[3].

Latent user interaction studied by using three component. First by analyzing characteristics of RenRen social graph & compare with other OSN. Second, focusing on latent interaction by describing log reconstruction algorithm which uses clock to merge visitors log. Finally build latent interaction graph from visitors log[4].

It studies the services in web based video sharing sites which includes uploading the user generated video, commenting & rating of video.[5]

In this peer to peer structure is generated. Improvement in multimedia content delivery in Youtube[6].

III. PROPOSED SYSTEM

Problem definition:- "Analysis of external links of videos using Association Rule Mining"

In a proposed method, find & analysis on no. of hits of video from external links with different categories, personalization of user & also find relevant video links with comparative analysis of Apriori algorithm and parallel FP-Growth algorithms.

IV. ARCHITECTURE

In this diagram we can see the overall picture of the system. There collected videos are presents. When universal Java Script engine which parse the java script pages and then it stores the external links. This process is called crawling. It takes inputs as the information like no.of views of videos and ages. When user upload his video he insert information like from which category it belongs to, link, name of the video.

Experimental datasets collected from two user generated content VOD sites, YouTube and Youku. YouTube is one of the largest UGC VOD sites in the world and at the time, and it accepts roughly 1.986 billion views every day.

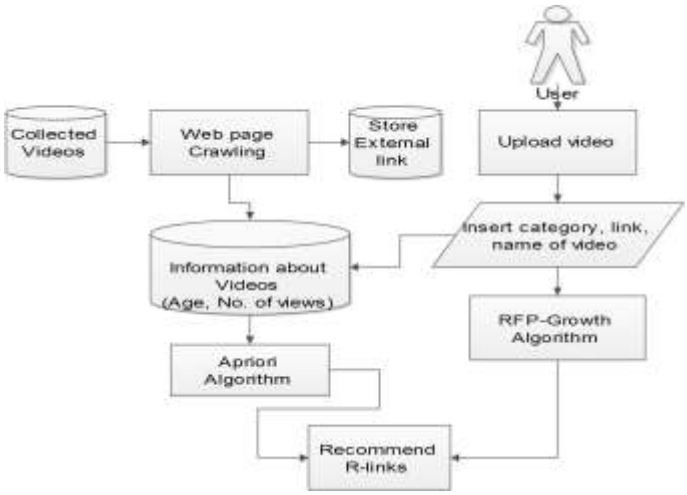


Fig: System Architecture

In China Youku is the most popular video site, so views comes through external links in Youku is relatively more than You tube.

V. MATHEMATICAL MODELING

Let system S

S={U,Vdatabase,C,EL,IL,RFP-Growth,V_Details}

U is no. of users{*U1, U2,U3,….*}

Vdatabase=No.of videos in database={*V1,V2,V3*}

V_Details={Upload_date, no of views, Related IL,EL,C}

Upload_date= to find age of video

EL={External Links of videos}

IL={Internal links of video}

No of views NV={*NV1,NV2…*}

C is category of video={*C1,C2,C3....*}

RFP Growth={*C,N,U,T,Q,H,N*}

H=Header table={*H1,H2*}

Q=node links={*Q1,Q2,....*}

T=Tree generated by link nodes

N=Nodes={*N1,N2,....*}

VI. IMPLEMENTATION STRATEGY

For implementation of this paper we used following technique & algorithm.

A. *Association Rule Mining*

Association rule mining are one of the major techniques of data mining and it is used to find the most common form of local pattern discovery in unsupervised learning systems. It act as a useful tool to finding correlations between items in large databases.

Algorithms of Association Rule Mining :

a) ***Apriori Algorithm:***

Apriori is an algorithm which finds the frequent itemsets by using candidate generation. It is a level wise complete search algorithm using anti-monotonicity of itemsets, if suppose itemset is generally not frequent one, any of its superset is also nota frequent. Let the set of frequent itemsets having size k & Lk and their candidates be Nk It iterates over the following three steps and extracts all the frequent itemsets:

1. Generate Nk+1 , candidates of frequent itemsets of size k +1, from the frequent itemsets of size k.

2. Scan the whole database and calculate the support of each items of frequent itemsets.

3. Add those itemsets that satisfies the minimum support requirement to Lk+1.

It is of two steps:

 ISSN: 2321-8169
4811 - 4814

- Join step: Generate Rk+1 , the initial candidates of frequent itemsets of size k+1 taking the union of the two frequent itemsets of size k, Pk and Qk that have the first k1.

- Prune step: We have to check if all the itemsets of size k in Rk-1 are frequent and generate N by removing those that do not pass this requirement from Rk+1.

b) *RFP*

RFP algorithm is improvement over FP growth algorithm for mining.

RFP algorithm is improvement over FP growth algorithm for mining.

1) Scan the database ,calculate support for each item & save it into header table having 3 fields like name,support,link.

2) Remove items whose support less than minimum support & arrange in descending order.

3) Scan database repeat over three steps:

3.1) remove non-frequent
3.2) sorting
3.3) After ordered itemset it is added to tree.

4) process each itemset of the header table from last one. Suppose current itemset is Q.

4.1) add Q into base itemset.

4.2) read all items from node to tree then save the support to sub header table.

4.3) Remove support which doesn't satisfy min sup.If items in BI & subheader are same or it is empty.

4.4) Read all the items from node Ni (i= 1, 2, ..., k) to the root of tree T.

4.5) sorting subheader table.

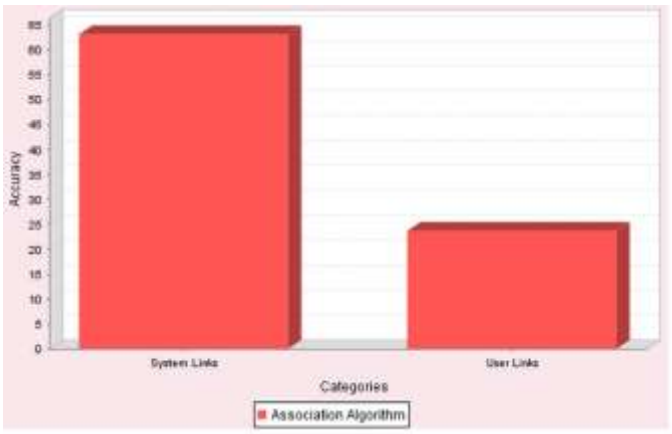
4.6) When ordered itemsets are added to the subtree subT, all the support number of the itemsets' corresponding nodes plus s then keep all the new nodes in the link of corresponding subhead table.

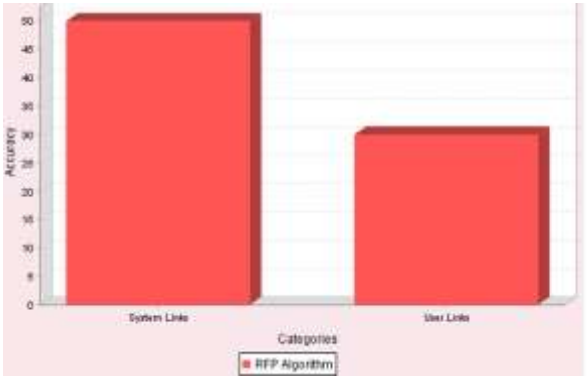
4.7) Process sub-header table subH from step 4 recursively;

5) Remove the current processing item of the current processing header table from base itemsets BI, then continue process next item in the current processing header table.

VII. RESULT ANALYSIS

The results of the algorithm are captured by comparing the two algorithm based Average no of links & total number of external links.





Here in graph we can see that in this algorithm we can see that total number of relevancy is generated by RFP algorithm is greater than Apriori algorithm. And as the view from RFP increases the relevancy is also increased.

VIII. CONCLUSION

The external links have an important role in the distribution of the videos. The external links are having different impacts on YouTube and Youku.In this we are going to find relevant video with the help of personalization. For implementing this two algorithms of Association Rule i.e Apriori and RFP growth are used. Both the algorithms which are selected were able to discover access pattern and user behaviours using support and confidence thresholds accurately. Memory requirement of the Apriori algorithm does not care about the number of transactions while the memory requirement of the RFP-growth algorithm increases significantly with the growth of the number of transactions. The main drawback of Apriori algorithm is that there is huge candidate set generation, especially if a large number of patterns and/or long patterns

exist. From this it can be concluded that RFP is behaves better than Apriori.This study can be extended by classification or clustering algorithm to predict future user request**s.**

REFERENCES

[1] A. Brodersen, S. Scellato, and M. Wattenhofer, \YouTube around the world: Geographic popularity of videos," inProc. ACM WWW'11, Lyon, France, Apr. 1620, 2011.

[2] 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,"inProc. ACM IMC'07, San Diego, CA, Oct. 24{26, 2007

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

[4] J. Jiang, C. Wilson, X. Wang, P. Huang, and W. Sha, "Understanding latent interactions in online social networks," in Proc. ACM IMC'10, Melbourne, Australia, Nov. 1–3, 2010.

[5] S. Mitra, M. Agrawal, A. Yadav, N. Carlsson, D. Eager, and A. Mahanti, "Characterizing web-based video sharing workload,"ACM .

[6] S.Scellato,C.Mascolo,M. Musolesi,and J. Crowcroft, "Track glob-ally, deliver locally: Improving content delivery networks by tracking geographic social cascades," inProc. ACM WWW'11, Hyderabad, India, Mar. 28–Apr. 1, 2011.