FACILITATING VIDEO SOCIAL MEDIA SEARCH USING SOCIAL-DRIVEN TAGS COMPUTING

Wei-Feng Tung and Yan-Jun Huang

Department of Information Management,
Fu-Jen Catholic University, Taipei, Taiwan

ABSTRACT

Online video search or stream live on social media has become tremendous widespread and speedy increased continuously in recent years. Most of the videos shared on social media are aimed at the more number of views from audiences. What and how many videos the users shared all around the world have created a great amount and varied videos and the other data into Internet cloud's database and even can be viewed as a kind of big data of digital contents. This research is to present how to implement a social-driven tags computing (SDT) which can be used to facilitate online video search on social media platforms.

KEYWORDS

SDT, Tags Computing, Video Search, Social Media, Social-Driven

1. INTRODUCTION

The video social media and social networks have widely and deeply used no matter on websites or on mobile phones Apps (i.e., YouTube, Instagram, Vine, Dubsmash, Snapchat…etc.) Social media have also tremendous changed the ways to communicate and share things to the other people. Furthermore, social media enable the users exchange, interact, and share the many resources on their communities of social networks. For example, video shared media enables their uses all over the world upload or search all of online videos' resources. As for the users, sponsors, and advertisers, how to promote the videos is their motivations and goals.

In order to increase the effective videos' search via the social media, there are some new methodologies that can facilitate the users to share and browse these video resources effectively. Online users can initiatively input one or more tags (keywords) when they upload their videos. Tagging is also to allow the users actively add one or more tags resources come from users' thoughts. The relevant tags can be determined from the existing tags' databases. Some of these tags (or keywords) might be added by the authors. These tags created just increase the possibility of being searched, and a way to share the authors' opinion toward the resource. Most of the time, these social tags are created according to the users' perceptions toward the resources instead of written by the scholars or authorities of the resource sorting system.

David C. Wyld et al. (Eds) : ITCS, SIP, CST, ARIA, NLP - 2017
pp. 01– 09, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.70301

Majority social media platforms allow any user to state tags. Thus, the users would judge the resources according to their personal experience. Sometimes, users just do not have any idea in mind, not sure which tags are suitable for the resources. Concerning folksonomy, it can be collected a group of users under cooperation and sharing condition in public, adding tags or marks to provide meaning to certain resource. In this research, a social-driven tags computing (SDT) can be represented and used to help the users for their tagging and further facilitate the video searching.

2. LITERATURE REVIEW

Social tagging is the practice of generating electronic tags by users rather than specialists as a way to classify and describe content. Comparing with the information based on scholars or experts, social-driven tags computing (SDT) is a kind of new tagging model, which is also a user-generated classification. The reason why social media include tagging function is to help the users classify their video resources, and the increase of spam tags would destroy the good will of SDT function. Hence, the paper still adopts and implements the SDT computing, and expects the improvements of tagging mechanism on social media.

2.1 Tagging

Most of tagging websites include bookmark, photos, index, video, and blog [1]. Basically, if there is no word on the resource such as photos and videos, what we need most importantly would be users or resource authors' tag for sorting the resources. Since resources like video and photos normally lack word descriptions, it would be too hard to classify tags for the users. If we add a great amount of tags from users or authors' perceptive it would give us a hand for resources indexing and or searching [2]. Even if the resources from blog are mainly formed with words, sometimes these blogs would cause the problems such as too much content or the meaning of words with diverse meaning.

Annotating tags can be defined as a tagging behavior likes keyword for describing the Internet resource. Basically, a definition of tags is similar to keyword indexing. Tagging also possesses the function of content resource classifications [3]. Tagging is the first level analysis, and classification is the second level analysis-paralysis activity [4]. It would be a magnificent task to form a sorting framework, and then reorganize the tags with the framework and model. In comparison with traditional sorting system, bookmark, tagging is relatively easy for users to learn and use, it would not increase the burden of cognitive, and it's easier for maintaining.

Generally speaking, SDT is the organizing model which combines the public's tags or keywords to form the main topic/theme for classification. Every Internet user has his or her own information management model, including personal bookmark, tags, index, email file document etc. Some are sorted with the set level or classification framework; some are sorted with the keywords that are qualified to be recognized, still other even without sorting in advance [5].

2.2 Folksonomy

Folksonomy is the combination of "Taxonomy" and "Folk," which means classification and people [6]. One of the meanings of Folksonomy is a group of users under cooperation and sharing condition in public, adding tags or marks to provide meaning to certain resource. It does not have concept of level, but it has the trait of clustering, meaning that once the resource is tagged more

and more times, it could create new definition to the resource and replace the definition laid by experts. Folksonomy does not request the people who classify the resource with professional knowledge, what's more, it encourage the users to sort the resource freely so that the loose classification structure could become convergent gradually and form the definition that could be accepted by the public and scholars.

This kind of sorting mechanism is called Folksonomy. Folksonomy's meaning is close to Social-driven tags computing (SDT). This concept was created by Thomas Vander [6]. Folksonomy is a different way of classification from the traditional systematic classification system. It is conducted by the public, which would come up with the tags or comments toward the resources that are closer to the opinions or feedbacks given by the users.

3. SOCIAL-DRIVEN TAGS COMPUTING

The research proposes a social-driven tags computing (SDT) which can provide online users an enhanced list of tags from the existing tags' database as well as video search. In terms of search engine optimization (SEO), majority online video web sites adopt various advanced or innovative recommendation technologies that can efficiently help their users to share their videos and tag their videos' metadata as shown in Figure one.

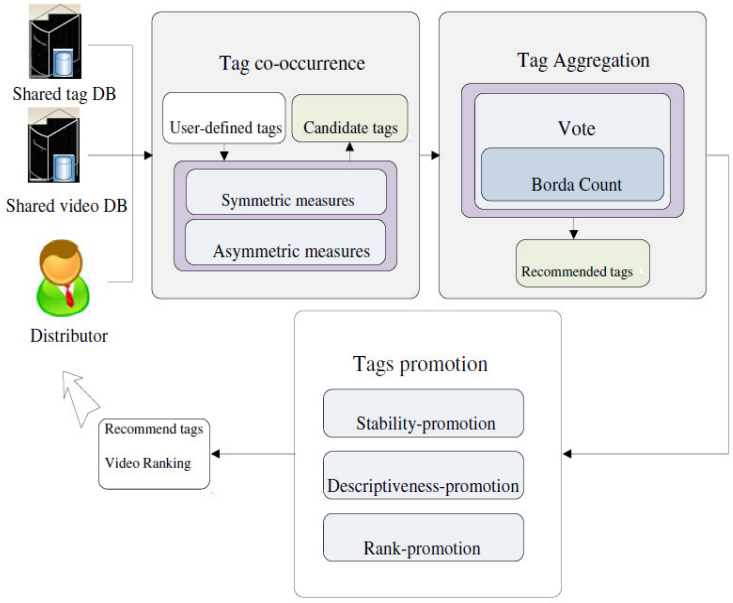

Figure 1. Social-driven tags computing (SDT) framework

3.1 Social-Driven Tags Computing

The first step of this research is to estimate a similarity measurement. Similarity in common use has twofold: symmetric measurement and asymmetric measurement.

Symmetric similarity measurement: Jaccard coefficient can be used to measure the co-occurrence value between tag t_i and tag t_j to measure the degree of similarity shown in Equation 1 [7].

$$J(t_i, t_j) := \frac{|t_i \cap t_j|}{|t_i \cup t_j|} \quad (1)$$

Jaccard coefficient $J(t_i, t_j)$ indicates that the interaction of tag t_i and tag t_j, divide the union of tag t_i and tagt_j. Jaccard coefficient is applied to measure the similarity of relative tags to determine the tag candidate in this VPA.

Asymmetric similarity measurement: The count of single tag can be normalized to further assess the tag co-occurrence value as follows in Equation 2.

$$P(t_j|t_i) := \frac{|t_i \cap t_j|}{|t_i|} \quad (2)$$

The probability of tag $t_i$and tag t_j represents simultaneously while tag t_i appears.

3.1.1 Tags Aggregation

The tagging candidates can be determined when the co-occurrence of tags was assessed. In the next procedure, the candidate tags can be integrated into a candidate tags list. The first step indicates that these candidate tags need to sort by aggregation using vote and sum. The second step indicates the filtered tags and recommended tag list can be generated through 'Borda count' [8].

While similarity measurements are done with calculation, the tagging aggregation function is the second procedure and the third promotion function for the determination of ranking objects. In the next voting process, the recommended tags can be decided by Borda count. The tags derived from the different candidate tags are compared to the other sets of tags. These selected tags can be voted 1 or 0 and further determined for recommended tags in the next process.

$$vote(u, c) = \begin{cases} 1 & if\ c \in C_u \\ 0 & otherwise \end{cases} \quad (3)$$

If the recommended tags that are selected from the candidate tags are determined, the scores of recommendation can be rated by the counts of voting (u, c) in the voting process.

$$score(c) := \sum_{u \in U} vote(u, c) \quad (4)$$

3.1.2 Tags Promotions

Most of tags are annotated in the shared tag archive; these tags are usually identified as the unstable tags for recommendation. On the contrary, some tags would be useful to describe the shared object more so than others. However, the tag promotion functions are threefold: stability-promotion, descriptiveness-promotion, and ranking-promotion.

3.2 Integrated Tags Computing

The tag prompt approach is to facilitate the further determinations of ranking scores from the candidates for recommended tag.

Stability-promotion: In order to promote those tags for which the statistics are more stable, the frequency of usages of tags can be measured to represent the levels of stability shown in Equation 5.

$$stability(u) \coloneqq \frac{k_s}{k_s - abs(k_s - \log(|u|))} \tag{5}$$

where |u| represents the frequency of tag u, k_s is a parameter for training.

Descriptiveness-promotion: If the descriptiveness-promotion is high frequency, the shared tags can increase the high frequency for the recommendation of shared objects shown in Equation 6.

$$descriptive(c) \coloneqq \frac{k_d}{k_d + abs(k_d - \log(|c|))} \tag{6}$$

where k_d is a pre-defined training parameter; c is one set of candidate tags.

Ranking-promotion: The co-occurrence provides a good evaluation that can estimate the relationships among the shared tags (u) and change to the ranking(r) for candidate tags ($c \in C_u$) shown in Equation 7.

$$rank(u, c) \coloneqq \frac{k_r}{k_r + (r - 1)} \tag{7}$$

where k_r is a damping parameter.

According to the three different promotion-functions, a holistic promotion value can be estimated by multiplication shown in Equation 8.

$$promotion(u, c) \coloneqq rank(u, c) \cdot stability(u) \cdot descriptive(c) \tag{8}$$

Based on the aggregation methods of Vote and Sum, the score can be computed by vote and promotion.

$$score(c) \coloneqq \sum_{u \in U} vote(u, c) \cdot promotion(u, c) \tag{9}$$

$$score(c) \coloneqq \sum_{u \in U} vote(u, c) \cdot rank(u, c) \cdot stability(u) \cdot descriptive(c) \tag{10}$$

Where score is the voting results; the three promotion functions use the multiplication of $rank\ (u, c)$,$satability\ (u)$, and $descrriptiv$e (c).

In terms of another mode, it can combine Sum and promotion function.

$$score(c) \coloneqq \sum_{u \in U} (P(c|u) \ \ , if \ c \in C_u) \cdot promotion(u, c) \tag{11}$$

$$score(c) \coloneqq \sum_{u \in U} (P(c|u) \ \ , if \ \in C_u) \cdot rank(u, c) \cdot stability(u) \cdot descriptive(c) \tag{12}$$

Where score (c) is the sum of voting and promotion functions, $promotion(u, c)$ is a multiplication of $rank(u, c)$, $stability(u), and\ descriptive(c)$.

Different combinations of vote, sum, promotion function, and no-promotion function can be used to focus on the different types of shared videos and tags archive [9].

In terms of the recommendation technology, the vote-promotion algorithm (VPA) estimates the ranking scores based on vote value, stability value, descriptive value, and rank value for the results of video-tag relationship prediction. VPA is capable of measuring the degrees of relevance in a numerous collection of tags from the shared video archives.

The algorithm of tagging computing can help the distributors predict a ranking list of recommended tags and videos based on the other relative tags. Figure 6 shows that the recommended tags can be analyzed and determined when users post the initial two tags. This distributor can obtain 6 recommended tags (i.e., Billie Jean, Michael Jackson, Singer, soul, live, and mj) if he posts two tags 'Thriller' and 'Moonwalk'.

3.3 System Process of SDT Computing

The proposed social-driven tags computing (SDT) adopts a 'Crawler' system to search for the relative tags on the video sharing websites. All the names and tags of the shared videos are stored in a video database shown in Figure 2 as resources for video query.

```
<title>AC/DC - Thunderstruck - YouTube</title><link rel="search" type="application/opensearchdescription+xml"

<meta name="description" content="Music video by AC/DC performing Thunderstruck. (C) 1991 J. Albert &amp;

<meta name="keywords" content="AC/DC, Thunderstruck, Epic, Pop">
```

Figure 2. Source codes of AC/DC video web pages on Youtube

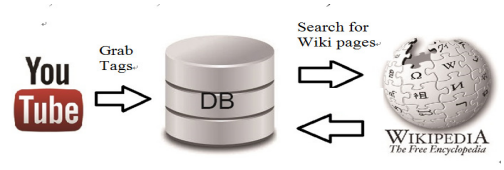

Figure 3. Tag resources from Wikipedia

Wikipedia can be used to refer to the determination process and further adjust the sequence of tags according to the relative tags or terms form searched tags estimations. For example, 'Thriller' tags can be changed its relative list of tags.

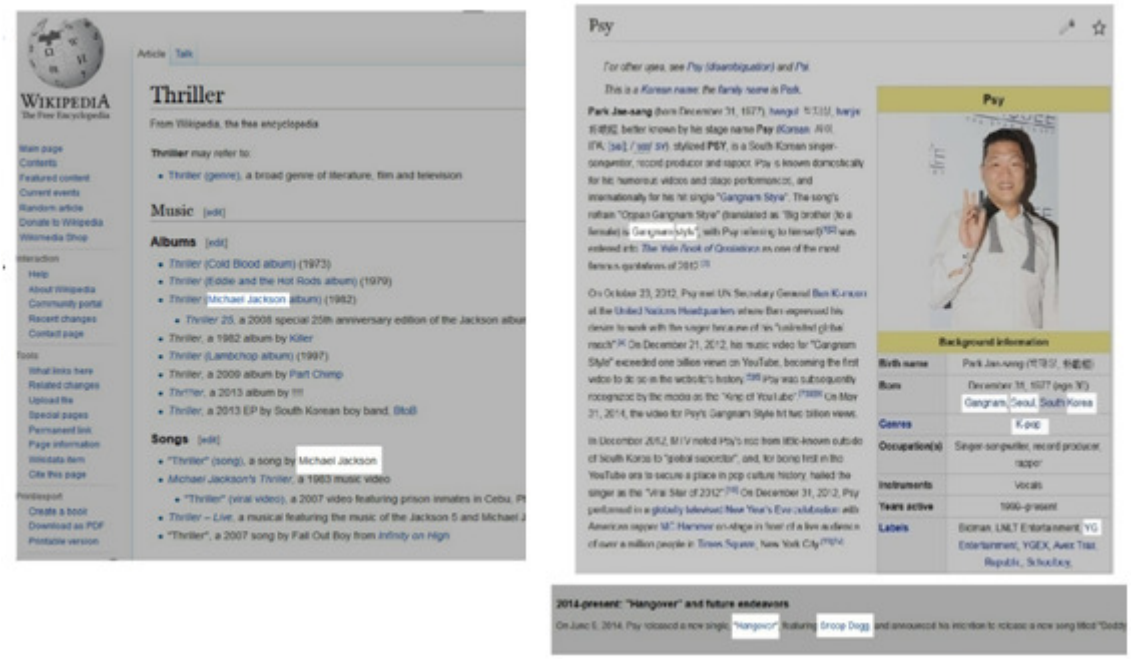

Figure 4. Wikipedia data facilitate the tag determination

When a user intends to upload a video and needs to provide the tags at the same time, the tags system developed by the research can then generate a list of recommended tags from the video database and wikipedia as shown in Figure 3 and 4. Determined by SDT, the weights are also represented for ranking (Fig. 5)

Figure 5. SDT Computing User Interface

SDT estimates the 'Jaccard coefficient' to calculate the co-occurrence values to provide the candidate tags based on the particular tags from the user queries as the follows figure 6.

Figure 6. Tags recommended for Video Search

To improve the efficiency ranking of search, more detailed tags should be given higher weight than general tags as shown in Table 1. To adjust the weights of tags by the computing of ranking, the promotion estimations can be facilitated for video search (Fig. 7)

Table 1. Tags Estimations by Promotion

Tags	Parachutes	Stories	Ghost	Magic	Paradise
Ranking Promotion	0.9523	1	0.9090	0.7820	0.8
Stable Promotion	0.6356	0.6356	0.6356	0.6356	0.6356
Describe Promotion	0.7153	0.6705	0.5670	0.7820	0.5916
Usage Frequency	4	31	58	19	49

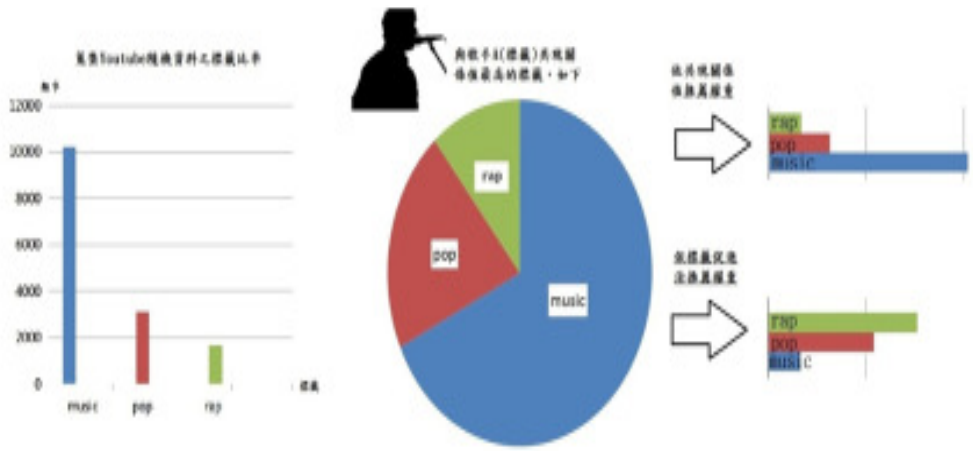

Figure 7. The weights of tags can be adjusted by SDT computing

4. CONCLUSION

Social-driven tags computing is to facilitate online video search and even sharing. As most of the videos shared on social media are aimed at the more number of views from audiences, the users (distributors) want to annotate some valuable tags. What and how many videos the users make decision by themselves, but this research can help the users to choice the other recommended tags for the specific video resources. The SDT methodology includes the co-occurrence estimations and tags voting as well as promotions like stability, descriptive, and rank. Those algorithms are able to determine the valuable tags according to the existing tags' databases in the social media.

REFERENCES

[1] Hammond, T., et al. , "Social bookmarking tool", D-Lib Magazine Vol. 11, no. 4, 2005.

[2] Trant, J. and Wyman, B., "Investigating social tagging and folksonomy in art museums with Steve.Museum", Proceedings of the Www Collaborative Web Tagging Workshop, 2006.

[3] Marlow, C., Naaman, M., and Boyd, D. "HT06, tagging paper, taxonomy, Flickr, academic article, to read", Conference on Hypertext. 2006.

[4] Sinha, R.(2005, September). A cognitive analysis of tagging. [online] Available: http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html.

[5] Jones, W., Dumais, S., and Bruce, H. "Once found, what then? A study of "keeping" behaviors in the personal use of Web information". Proceedings of the American Society for Information Science & Technology, p.p. 391-402. 2002.

[6] Thomas Vander Wal. Explaining and showing broad and narrow folksonomies. 2005.

[7] Loet Leydesdorff, "Regional Development in the Knowledge-Based Economy: The Construction of Advantage". Journal of Technology Transfer. Special Issue, pp. 1–15, 2006.

[8] Singurbjornsson, B. and Zwol, R. V., "Flickr tag recommendation based on collective knowledge", Proceeding of the 17th International Conference on World Wide Web, 2008, pp.327–336.

[9] Christian Wartena, Rogier Brussee, and Martin Wibbels, "Using Tag Co-occurrence for Recommendation", International Conference on Intelligent Systems Design & Applications, 2009, pp.273-278.