Concept Discovery in Youtube.com using Factorization Method

Janice Kwan-Wai Leung

Abstract

Social media are not limited to text but also multimedia. Dailymotion, YouTube, and MySpace are examples of successful sites which allow users to share videos and interact among themselves. Due to the huge amount of videos, categorizing videos with similar contents can help users to search videos more efficiently. Unlike the traditional approach to group videos into some predefined categories, we propose to facilitate video searching with clustering from comment-based matrix factorization and to improve indexing via the generation of new concept words. Factorized component entropies are introduced for handling the difficult problem of vocabulary construction for concept discovery in social media. Since the categorization is learnt from users feedback, it can accurately represent the user sentiment on the videos. Experiments conducted by using empirical data collected from YouTube shows the effectiveness of our proposed methodologies.

1 Introduction

In the Web 2.0 era, people can interact effectively in the Internet instead of just retrieve data. Social networks such as forums, blogs, video sharing sites are examples of applications. Social networks like Facebook [3], Bebo[1], Flickr [4] are blooming with user generated contents which can be in forms outside text such as images or videos.

Recent years online video sharing systems are burgeoning. In video sharing sites, users can upload and share videos with other users. YouTube [6] is one of the most successful and fast-growing systems. In YouTube, users can share their videos in various categories. Among these video categories, music is one of the most popular one and the number of music videos overly excess that of other categories [9] [19]. Users are not only allowed to upload videos but tag videos but leave comments on them as well. With more than 65,000 new videos being uploaded every day and 100 million video views daily, YouTube becomes a representative community among video sharing sites [7].

Due to the incredible growth of video sharing sites, video searching is no longer a easy task and more effort should be paid by users to search their desire videos from the entire video collection. To address this problem, grouping videos with similar contents together and indexing are necessary. As such, information about video content is needed for the objective mentioned above. However, it is an even challenging problem to find out accurate information about the uploaded videos. Currently, videos on YouTube are only coarsely grouped into some predefined high level categories (e.g. music, entertainment, sports, etc) in which category of a video is just decided by a single user who put up the video. Under this policy of predefining categories, videos in a single category still span through a wide range of varieties. For example, in the music category, we may find music from various countries or with different musical styles. Though some other video sharing sites, such as DailyMotion [2] and MySpace [5], have a lower level of category for music videos, the categorization just follow the basic music genre. However, people attentions to music are not limited to these simple genre. Furthermore, the predefined categories maybe too subjective to capture the real attracted issue of singers to the majority of users since they are only defined by a small group of people. Finally, the current categories on YouTube are fixed and it is hard to add/remove categories too often. As time goes by, some categories may become obsolete and some new topics may be missing from the categories.

These observations motivate us to explore a new way of video categorization for facilitating video search. In this work, we propose a novel commentary-based clustering technique by utilizing user comments for achieving this goal. Unlike the traditional approaches of predefining some categories by human, our categorization is learnt from user comments. The advantage of our proposed approach is three-fold. First, our approach can capture public attentions more accurately and fairly than that of the predefined categories approach as we have taken the user opinions into consideration. In other words, the resulting categories are contributed by public users rather than a small group of people; Second, since user attentions can be changed from time to time, the categories of our method can be changed dynamically according to the recent comments by users; Finally, as users comments are in the form of natural language, users can describe their opinions in details with rich text. There-

fore, by commentary-based clustering, we can obtain clusters which represent fine-grained level ideas of videos.

In the literature, various clustering techniques have been proposed for video categorization [16] [26]. However, this type of techniques did not take user opinion into consideration and thus the clustering results do not capture public interested issues.

Apart from predefined categories, YouTube also provides tagging to assist video searching service. Indexing videos with some words describing the videos should theoretically be helpful in the context of video understanding while users would not have any idea before viewing a video. Nevertheless, the tags are usually too loose and not structural which are hardly to give enough description of videos.

Some researchers have proposed to use the user tags on videos for clustering [17] [15]. Though user tags can somehow reflect user feelings on videos, tags are, in many cases, too brief to represent the complex ideas of users and thus the resulting clusters may only carry high-level concepts. Another stream of works which use commonly fetched objects of users for clustering [11] suffer similar shortcoming of neglecting object content. In [25], they proposed to adopt a multi-modal approach for video categorization. However, their work required lots of human efforts to first identified different categories from a large amount of videos.

We want to remark that although commentary-based clustering can theoretically obtain more fine-grained level clusters, it is much more technically challenging than that of tag-based clustering. The reason is that user comments are usually in the form of natural language and thus preprocessing is necessary for us to clean up the noisy data before using them for clustering.

The rest of the paper is organized as follows. Section 2 discusses previous works in the context of social network mining. Section 3 explains our proposed approach for video categorization in video sharing sites. Section 4 briefly introduces our web crawler. Section 5 presents the details of pre-processing of the raw data grabbed by our crawler. Section 6 describes our video clustering algorithm. Section 7 presents and discusses our experimental results. Section 8 concludes the paper.

2 Related Works

Since the late eighties, data mining has became a hot research field. Due to the advancing development of technologies, there is an increasing number of applications involving large amount of multimedia. For this reason, researches in the field of data mining are not limited to text mining but multimedia mining. Qsmar R. Zaine et al. [26] developed a multimedia data mining system prototype, MulitMediaMiner, for analyzing multimedia data. They proposed modules to classify and cluster images and videos based on the multimedia features, Internet domain of pares referencing the image of video, and HTML tags in the web pages. The multimedia features used include size of image or videos, width and height of frams, date on which the image or video was created, etc. S. Kotsiantis et al. [16] presented a work to discover relationships between multimedia objects based on the features of a multimedia document. In their work, features of videos such as color or grayscale histograms, pixel information, are used for mining the content of videos.

Motivated by the bloom of social networks, plenty of works have been done involving the study or analysis of online social networks. Different approaches are proposed to discover user interests and communities in social networks. Tag-based approach is one of the invented methods. In [17], Xin Li et al. developed a system to found common user interests, and clustered users and their saved URLs by different interest topics. They used the dataset from a URLs bookmarking and sharing site, del.icio.us. User interests discovery, and user and URLs clustering were done by using the tags users used to annotate the content of URLs. Another approach introduced to study user interests is user-centric which detects user interests based on the social connection among users. M. F. Schwartz et al. [20] discover people's interests and expertise by analyzing the social connections between people. A system, Vizster [14], was designed and developed to visualize online social networks. The job of clustering networks into communities was included in the system. For this task, Jeffrey and Danah identified group structures based on linkage. Except the use of sole tag-based or user-centric approaches, there are works done with a hybrid approach by combing the two methods. In [15], user interests in del.icio.us are modeled using the hybrid approach. Users are able to make friends with others to form social ties in the URLs sharing network. Julia Stoyanovich et al. examined user interests by utilizing both the social ties and tags users used to annotate content of URLs. Some researchers proposed the object-centric approach for social interests detection. In this approach, user interests are determined by the analysis of commonly fetched objects in social communities. Figuring out common interests is also a useful task in peer-to-peer networks since shareds interests facilitate the content locating of desire objects. Guo et al. [11] and K. Sripanidkulchai [22] presented in their works the algorithms of examining shared interests based on the common objects which users requested and fetched in peer-to-peer systems.

3 Public Attention Based Video Concept Discovery and Categorization for Video Searching

With the ceaseless growth of media content, it is increasingly a tense problem for video searching. It is usual that users hardly find their desire videos from the immense amount of videos. There are two main directions to ease the process of video searching, one is enhancing the text-based search engine whilst the other one is designing a better directory. In this paper, we focus on the former approach.

Though many video sharing sites allowed tagging function for users to use tags to annotate videos during the upload process, it is very common for user to tag videos by some high level wordings. As such, tags are usually too brief for other users to locate the videos by using the text-based search engine. In our method, as user comments usually describe the videos in details, we can use them for video clustering to obtain fine-grained categories. By identifying the concept words for each categories, we can use them as latent tags for the corresponding categories in order to facilitate the video searching process.

In music domain, music videos in sharing systems are always categorized according to their types of musical sounds (e.g. pop, metal, country, etc.) under the music genre. However, except music styles, people may have many different attitudes and preferences (e.g. appearance of singers, event of performance, age of songs, etc) towards music in different regions. Therefore, to categorize music based on publicly interested issues, music genre is not a good categorical construct for video searching.

Our aim is to find a categorization where videos in each video group are representing a popular topic of interest and improve index with the in-depth concept of videos. In our algorithm, public attentions are modeled and video concepts are discovered by clustering videos into groups with the utilization of user-left comments.

Previously, computer scientists have tried many ways to find user interests. Tags are very popular to help in this context [17]. However, in a previous study of tagging in Youtube, it has been observed that many tags could not enhance the description of video as a result of system constraints [10].

Several disadvantages would be raised in this manner. Tags on a video are manually given by the one who uploads the video, thus the tags are just expressing a single user's feeling about the video. A study of content interactions in Youtube shows that tagging is unreliable as a result of self-promotion, anti-social behavior as well as other forms of content pollution [8]. Therefore, tags on a video would have a strong bias and are not fair enough to exactly describe what the video is actually about. Furthermore, single-user given tags are definitely not representative of public feelings about the video. To address the sparsity and ambiguity of tagging, folksonomy search has been suggested [18] to improve existing tags in video. However, such systems still depends on a set of content category tag which is self found in youtube.com.

Moreover, videos are often tagged with a small number of words. As such, often fails to give enough description on the video. Though there is a previous work classifying videos from youtube.com by using the tags, the reported average number of tags per video is just 8 to 9 which is far fewer than the amount of comments per video [21]. Therefore, tags are insufficient to provide detailed information about videos. Another study of tagging across four major social media websites has shown that only 0.58% of tags in youtube.com belongs to the content category. Such percentage is the lowest among the four major social media websites of study [13]. In order words, only a very small amount of tag can identify the content category of the video in youtube.com. Since comments can be given by any users on any videos as feedbacks, they express different users thoughts about a video. Thus, containing more in-depth information about the videos. Also, by allowing every user to leave feedbacks, the number of comments on a video are usually much more than that of tags. Hence, utilizing comments instead of tags to find out the attracted issues can solve the above difficulties.

In a study of video search in youtube.com, it is found that search services are critical to social video websites but users often cannot contribute to the search service [12]. In our proposed work, such problem can be addressed by involving the user-left comments to enhance video searching.

Mentioned above, though tagging is popular be used as an assistant in video sharing sites, it is yet far from perfect for video searching. Our proposed work is aimed to supplement the tagging technique to achieve the goal of providing a better video searching service for users.

Beside tag-based, some researchers proposed the content-based approach to categorize videos [23]. Using video content as categorizing materials can group similar videos together according to their actual content. Nevertheless, video content itself only provide objective information about the videos but nothing about users' idea. Consequently, this approach fails to group videos according to public attentions. In contrast, user-left comments include users' view about the videos. Therefore, comments can, undoubtedly, be used to categorize videos based on public attentions.

Video features can also be used to achieve the goal of videos clustering [16]. Video features, however, are hard to be extracted automatically. Due to the limitation of human resources, automatic information retrieval from mass amount of data is preferred. Also, using video features to cluster videos suffers the same shortcomings of content-

based as well. Because of information retrieval dealing with text is much easier than video features extraction, and comments, in addition to video content, provide users' views on videos, user-left comments are significant for clustering videos.

4 Dataset collection

YouTube is a video sharing platform on which users can upload their own videos for sharing purpose. Along with each video, a short description can be entered by the uploading user and tags as well. Apart from the video uploading user, other registered users can also contribute to the video surrounding text by leaving comments on the video. In this paper, we focused on the user comments of videos of Hong Kong singers in YouTube and did a comparison between comments and tags.

We first defined a set of 102 Hong Kong singer/group names. Given the set of singer/group names, we developed a crawler to firstly visit the YouTube web site and automatically searches from the site the related videos based on video titles and video descriptions. From the resulting videos, the crawler saves the URL of each videos for further process. For the convenience of gathering user comments, the crawler transforms the fetched URLs to links which link to the pages of "all comments" mode of corresponding videos. With all the transformed video URLs, for each link, the crawler is able to scrape the video web page and grab the video title, all the user comments and the user names of who left comments on the video.

In the data set acquired by our crawler, 19305 videos are grabbed with 102 singers and 7271 users involved.

5 Data Pre-processing

To ease the process of video searching by discovering the public attentions and categorizing videos, larger amount of data is required from video sharing sites. However, just the large-sized collection of text-formatted raw data is not applicable for further processing. Large-sized dataset always need to undergo data pre-processing in the field of data mining. Here is no exception in our algorithm. After crawling YouTube, the mass data need to be pre-processed before performing video clustering.

Here are two steps of data pre-processing involved in our introduced algorithm,

1) Data Cleaning
2) Text Matrix Generation

5.1 Data Cleaning

As the comments left on YouTube videos are written in natural languages which consist lots of non-informative words, such as "thank" , "you", etc, text processing with such materials must be caution. To avoid resulting a poor clustering, data cleaning is necessary for handling the noisy words.

In natural languages, there are many words that are not informative for clustering. These words would make the entire dataset very noisy. Applying a stoplist is one of the ways to clean up these words. Since some words are obviously not informative, it is easy to define a stoplist of noise. With a predefined stoplist, non-informative or distractive words can be strained from the dataset. After removing all the useless words by the stoplist, the dataset is then passed to the process of matrix generation.

5.2 Text Matrix Generation

Text-formatted data is not easy for further processing, it is more convenient to transform the data from text to matrix representation beforehand.

For example, the dataset can be represented by matrix A of size $n \times m$ where n is number of videos in the dataset and m equals to number of unique case-insensitive words in the dataset. In A, each row is a vector of video words and element $a_{i,j}$ is the frequency count of word j occurs in comments left on video i.

To transform the textual data into a more easy-computed text matrix, a dictionary is firstly built with the case-insensitive words in all the comments in the dataset. As comments are all in texts, linguistically, there exist many meaningless words in comments. These meaningless words, e.g. "is", "am", "the", "a", always occur in an extremely high frequency. Therefore, words occur in frequency exceeding a threshold should be discarded. On the other hand, words that seldom occur are probably not the important ones, so words with few occurrence should also be neglected. Therefore, we set an upper bound and a lower bound for word occurring frequency. All the words with frequency less than the lower bound or larger than the upper bound are filtered out. After filtering all the meaningless words, dictionary can then be built and matrix can be generated as well.

6 Video Processing via Clustering

In order to facilitate the video searching process, finding fine-grained video concepts and constructing a video category based on public attentions are crucial as there is no way to match a video with the desired ones without a deep understand of video content and people do searching with their interests in the usual practice.

As video comments left by users provide opinions about the video or singers in the video, some words in the comments are actually describing the fine-grained level concept

of videos. Therefore we can find video concepts analyzing the video comments. With the concept words discovered from comments, video indexing can be improved by incorporating those concept words. Hence, facilitating video searching and make it be done in a more accurate manner.

With the reason that public attentions are reflected from the comments users left on videos, grouping similarly commented videos together is a possible way to provide a good video categorization. Since the objective of clustering is to distinguish substantial amount of data and group similar objects together, clustering is an adequate algorithm for constructing a video category that can guide user to his/her desire videos.

Figure 1 shows the procedures of finding video concepts, discovering public attentions to Hong Kong singers and categorizing Hong Kong singer videos from YouTube.

6.1 Video Clustering and Concept Discovery

For our purpose of building a good video category and learn the video concept for easier video searching, Non-negative Matrix Factorization (NMF) is the chosen clustering algorithm [24]. We propose to apply NMF for clustering based on three reasons. First of all, NMF is a bi-clustering method. With a bi-clustering algorithm, comment words and videos can be clustered simultaneously. Thus, the main characteristics of video groups can be drawn while grouping videos with similar user views together. Additionally, NMF does not provide an absolute assignment of videos to groups. Absolute assignment clustering algorithms are not suitable for singer video clustering. In practice, a video can belongs to multiple groups. For example, a classic music video can be performed by a singer who is passed away. The video is said to be in both "classic" group and "died singer" group. As NMF calculates possibility coefficients of each video to different groups, a single video can be assigned videos to multiple groups. Finally, NMF is effective for clustering. Since we need to cluster a large amount of data, effectiveness is one of the concerns. An effective low-dimensional linear factor model is desired.

Comments on a video often capture users feelings about the video or describe the video. Videos are clustered into the same group if they bear comments with similar contents. Similar videos, therefore, can be grouped together and with their characteristics be revealed as publicly attracted ones.

Let A be the $n \times m$ video-word matrix generated in the process of data pre-processing, where n and m are the number videos and number of words in dictionary respectively. As all the elements in A are the occurrence counts of words in documents, they are greater or equal to zero. This makes matrix A a non-negative matrix.

Since the importance of a term to a document can be reflected by it's number of appearance, the well-known keyword measure in Information Retrieval $tf - idf$ is adopted for extracting important words. Within the dataset, all the comments of a video is aggregated and considered as a document. Importance of term i in document j is $w_{i,j}$ which is computed by using $tf_{i,j}$ (term frequency of term i in document j) and idf_i (inverse document frequency of term i). Terms that are important to a document are expected to appear many times in the document. For this reason, the term frequency is used to measure the normalized frequency of a term in a document. Suppose there are t distinct terms in document j, $tf_{i,j}$ can be computed as,

$$tf_{i,j} = \frac{f_{i,j}}{\sqrt{\sum_{k=1}^{t} f_{k,j}^2}} \tag{1}$$

where $f_{i,j}$ is the number of times that term i appears in document j. As words appear in many documents are not useful for distinguishing documents, a measure idf is used to scale down the importance of these widely-used terms. The inverse document frequency of term i is defined as,

$$idf_i = log\frac{N}{n_i} \tag{2}$$

where N is the total number documents in the dataset, and n_i is number of documents that containing term i.

After computing the term frequency and inverse document frequency, the importance weight of a term i in document j is defined as the combination of $tf_{i,j}$ and idf_i,

$$w_{i,j} = tf_{i,j} \times idf_i \tag{3}$$

The greater the weighting, the more the importance is the term to the respecting document.

From matrix A, a non-negative matrix X can be produced by calculating the importance weights. Each element in X is defined as,

$$x_{j,i} = w_{i,j} = \frac{a_{i,j}}{\sqrt{\sum_{k=1}^{t} a_{k,j}^2}} \times log\frac{N}{n_i} \tag{4}$$

By fitting a k-factor model to matrix X, where k equals to number of groups to be obtained, X is decomposed into two non-negative matrices W and H, such that $X = WH + U$. After matrix decomposition, W is in size of $n \times k$ and H is in size of $k \times m$.

Our objective is to find W and H such that $X \approx WH$. By iteratively updating W and H, we can obtain W and H by minimizing the following function,

$$F(W, H) = ||X - WH||^2 \tag{5}$$

with respect to W and H and subject to constraints that $W, H \geq 0$.

Figure 2 shows the decomposition of video dataset matrix. From the resulting matrices, relationships between

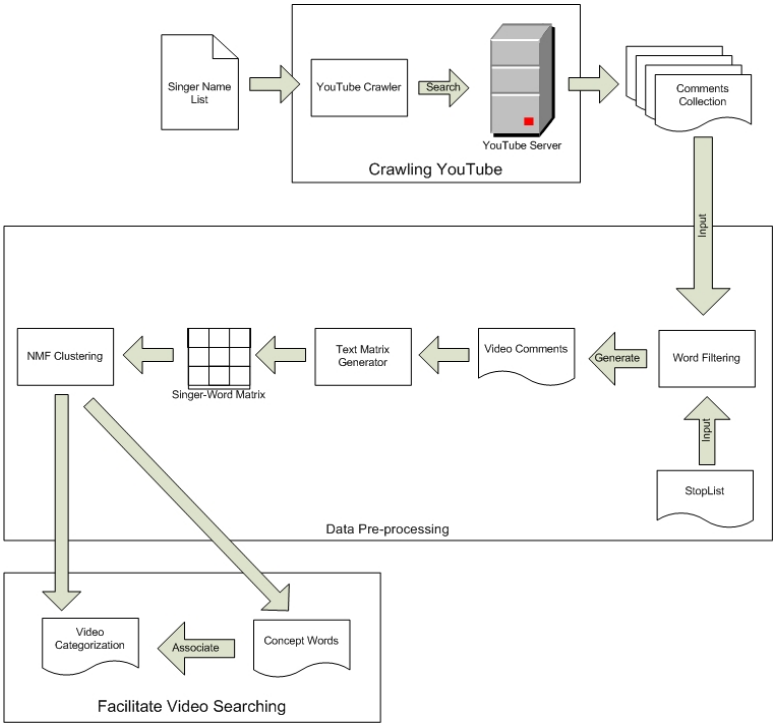

Figure 1. Video concept discovery and video categorization of Hong Kong singer videos in YouTube .

words, videos and clusters are revealed. Matrix W shows the relationships between videos and different clusters, whilst H clarifies the relationships between words and clusters. In W, value held in $w_{n,k}$ is the coefficient indicated how likely video n belongs to cluster k. To fit the purpose of our research, we have refined the method of group assigning in NMF. The original application of NMF algorithm assigns an object to a group in a maximum coefficient approach. However, in our method, video n is treated to be in group k if $w_{n,k}$ has the a value greater than a threshold β_k within vector n in W, where the value of threshold β_k is data dependent. The threshold should be chosen in a coefficient distribution depending manner. Videos can then be grouped into clusters based on their similarities. We define the set of clusters for video V_n that it belongs to as,

$$C_n = \{k \in K \mid \forall\, W_{n,k} > \beta_k\} \tag{6}$$

where K is set of all clusters.

Matrix H provides the information about the characteristics of the video groups. Concept words of a cluster can be found with H as $h_{k,m}$ is the coefficient of the term m belongs to cluster k. For each cluster, the top 10 words, with respect to the term-cluster coefficient, are considered to be the concept words for the cluster. Which the words states the properties of a group of videos and gives an in-depth

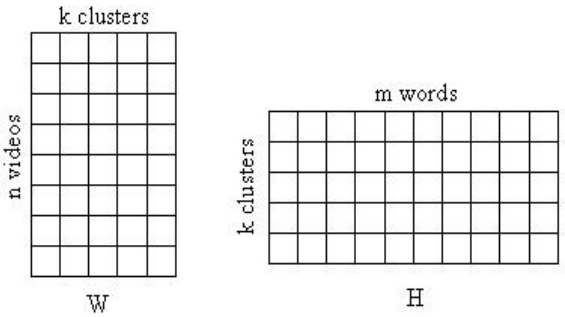

Figure 2. NMF decomposition for video clustering.

description for the videos. Enhancing video index by incorporating the discovered concept words can consequently improve users video searching experience.

6.2 Factorized Component Entropy Measures for Vocabulary Construction

While matrix factorization methods and latent Dirichlet methods have often been successful applied to process news articles and technical papers, applications of such algo-

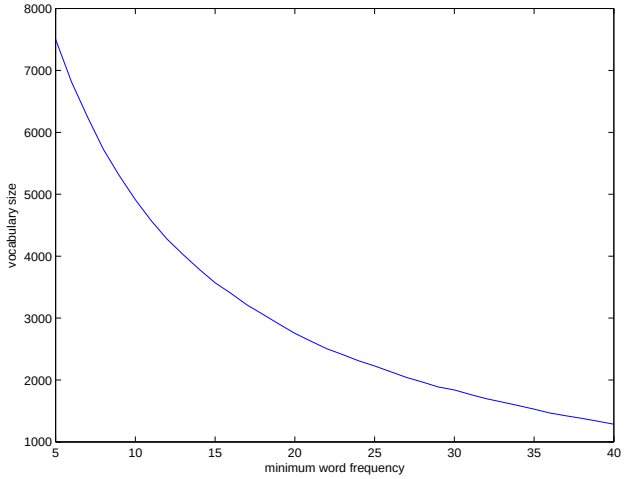

Figure 3. Vocabulary size.

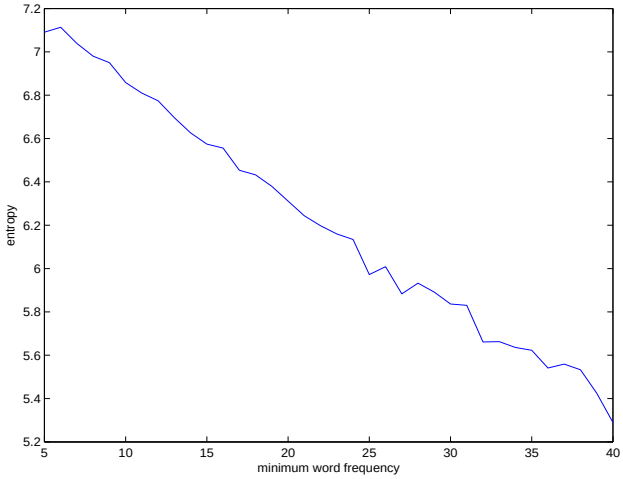

Figure 4. Word-concept entropy.

rithms to short and terse statements in commentary pose significant difficulties. Misspellings and the very short length of the commentary are often the norms in comments in youtube.com. We propose the use of factorized component entropy as a measure to construct good vocabulary for analyzing noisy commentary.

Figure 3 shows size of the vocabulary as a function of the global minimum word frequency where we can see a sharp drop in the size of vocabulary when the global word frequency is increased.

The two matrices W and H generated from factorization have the effect of indicating the cluster membership. The cluster membership c_i of the $i-$th concept is simply given by

$$c_i = \arg\max_j W_{ij},$$

where j is the concept label. To evaluate how the words are distributed among the different concepts, we can compute the word-concept entropy of the j-th concept using the following formula,

$$Ef_j = -\sum_i (H_{ij}/\sum_i H_{ij}) \log(H_{ij}/\sum_i H_{ij})). \quad (7)$$

A smaller word-concept entropy implies that the words in the features have coefficients in H that is distributed across a smaller number of features and is thus more favorable. A large concept entropy implies that the words have coefficients evenly distributed across the different concepts and thus cannot be clearly differentiated.

Figure 4 shows the word-concept entropy as a function of the global word frequency. As the global word frequency increase, and the size of vocabulary decreases which leads to a reduction in word-concept entropy.

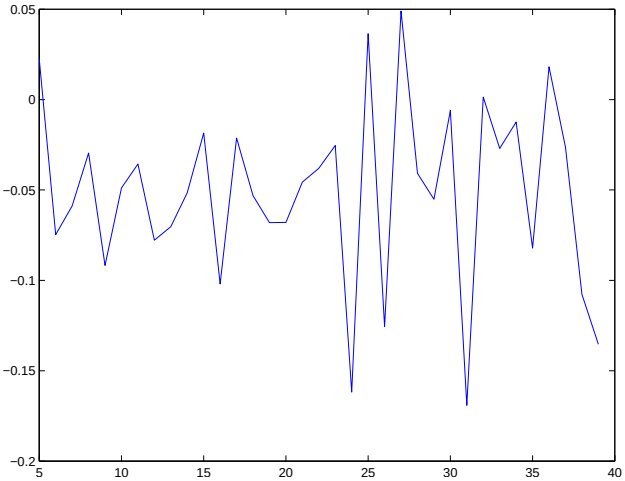

Figure 5. Derivative of word-concept entropy.

By taking the discrete derivative of the entropy, we can measure the change in entropy where the large drop in entropy represents the suitable size for vocabulary construction. Figure 5 shows the derivative of the word-concept entropy.

Similarly, we can also define the video-concept entropy which represents how well video commentary are grouped together using the following video-concept entropy formula,

$$Es_i = -\sum_j (W_{ij}/\sum_j W_{ij}) \log(W_{ij}/\sum_j W_{ij})), \quad (8)$$

where Es_i is the video-concept entropy of the j-th video commentary.

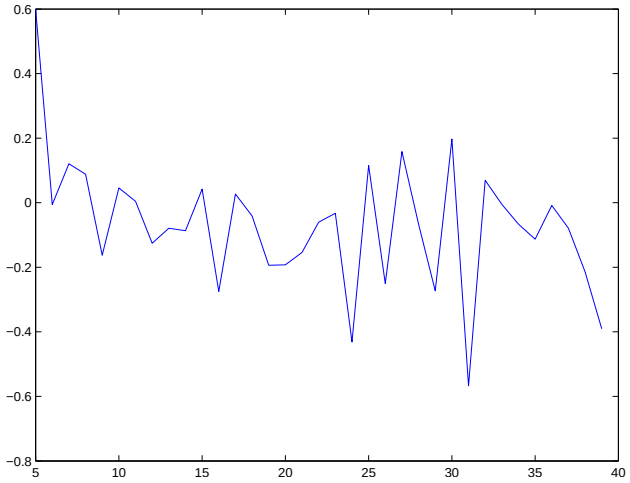

Figure 6. Derivative of joint entropy.

In the end, the joint entropy can be obtained by multiplying the video-concept entropy, word-concept entropy and the entropies similarly obtained by taking the transpose of W and H. The derivative of the joint entropy is shown in Figure 6, where the noises are further suppressed.

7 Experimental Evaluation

A proof-of-concept experiment was done to with videos in Hong Kong regional music domain. An Intel(R) Core(TM)2 Quad 2.40GHz PC with 4GB RAM was used to conduct our experiment. Our web crawler was implemented in VC++ and the core algorithm was implemented in Matlab.

7.1 Empirical Setting

As the videos were grabbed by searching from the YouTube site with predefined list of singer names, there are possibilities that some videos are grabbed more than one time. For those videos preformed by more than one singer, as long as there are more than one singer names annotated in the video title, the video will be collected in times equals to the number of hits the predefined singer name hits the video title. To achieve a more accurate clustering result, duplicated videos are removed from the dataset.

In comments, users are used to mention the singer names when they are commenting on him/her. This will make the singer names dominate in every group of concept words. However, it is not conspicuous enough to reveal detailed concept of videos by singer names. Therefore, in our experiment, we add singer names to the stoplist as well.

Furthermore, some videos are less popular or just been uploaded for a short time that only have a few comments. These videos which have relatively few words are non-informative for video clustering. Videos with commentary words less than the threshold discovered in earlier section are removed.

The videos are clustered into k groups with the clustering algorithm discussed in section 7, where k is experimentally set as 20. The experiment was done twice, once with threshold β_i regarding cluster i to be mean coefficient of all videos,

$$\beta_i = meanCoef_i = \frac{\sum_{j=1}^{n} w_{j,i}}{n} \tag{9}$$

To compensate the poor performance caused by the extremely uneven distribution of coefficient, we chose the threshold to be mean coefficient plus standard deviation of all videos for the second experiment. β_i regarding cluster i is defined as,

$$\begin{aligned} \beta_i &= meanSdCoef_i \\ &= \frac{\sum_{j=1}^{n} w_{j,i}}{n} + \sqrt{\frac{1}{n}\sum_{j=1}^{n}(w_{j,i} - \frac{\sum_{j=1}^{n} w_{j,i}}{n})^2} \end{aligned} \tag{10}$$

where n is total number of videos being clustered.

7.2 Video Categories and Concepts

Since video clustering is a complete clustering analysis, publicly attracted music categories in Hong Kong can be found by clustering the videos. We deployed NMF as our clustering method. Applied the clustering algorithm to the video dataset in the way discussed in Section 6.1, with the experimentally chosen number of cluster of 20, videos were clustered into groups based on the words in their comments. The mean coefficient of videos to a cluster is set as the threshold. Videos with coefficient higher than the threshold of a cluster are said to be in that cluster. Under this strategy, videos can belong to several clusters as they may have multiple characteristics. Table 1 shows the discovered categories and concepts from our dataset.

Unlike the generic music video categorization of some famous video sharing sites, such as DailyMotion divides music videos into eight classes (Pop, Rock, Rap, R&B, Jazz, Metal, Covers, and Electros), we categorized videos of local singers into twenty classes which are far more specific.

From our clustering result, we noticed that videos of singers are not only limited to general music videos, but also funny clips, award presentations, commercial advertisements as well as event promotion clips. Looking at the music videos alone, by clustering users' comments, we

Group	Concept Words
1	beautiful lyrics melody
2	female makeup dress
3	cute pretty handsome
4	sex photos scandal
5	funny hilarious laughing
6	rap raps hip
7	movie film story
8	cantonese mandarin language
9	commercial pepsi coke
10	piano piece ear grade
11	japanese japan korean
12	china olympic games
13	old classic memories
14	dance dancer moves
15	guitar band rock
16	award tvb gold
17	english chinese accent
18	sad legend died
19	together couple two
20	voice pretty talent

Table 1. Latent video categories discovered in Hong Kong music video domain from YouTube

Group	Top three cluster representative words
1	sin story her
2	ltd invisible target
3	木紋 如沾 數著
4	special 鐘泊桐 characters
5	我的第 hkpca awards
6	andrew yun fung
7	family food sheh
8	actors chin stephen
9	quot buenos zero
10	惺士山工 小南版 repeat
11	始終有 這地球 寺唱
12	bigboy2000 blogspot search
13	chi stephen derek
14	莫文蔚來囉 lollipop terry
15	label gold koon
16	takes goes 戀情告急
17	xuite daily blog
18	bird carina kar
19	lap jennifer wealthy
20	mahjong tak spirit

Table 2. Cluster representative words extracted from video meta data

found that people's attitude towards Hong Kong music are not only target on the music styles. There are also other features of music which people are interested in, like languages, age of music, music instruments, type of singers, singer's voice, composition goodness, etc.

Furthermore, categorize singer videos with the proposed clustering algorithm, people can identify dance-oriented videos (Group 14), cross-culture produced music (Group 11) or even movie theme songs (Group 7) easily. Other than simply categorizing singer video clips, some up-to-date news in the local music circle, like scandals (Group 4), can also be found.

Tags are popularly investigated in the contest of topic detection. To compare the effect of concept finding by comment with tags, an experiment was conducted with the same setting but video meta data as the dataset. Video meta data are all the video surrounding tests including title, tags and description. Contrastingly, representative words of clusters extracted from video meta data cannot bring any idea of the video groups. Table 2 lists the top three representative words of each resulting video group derived from meta data. From the words listed in the table, nothing about the video

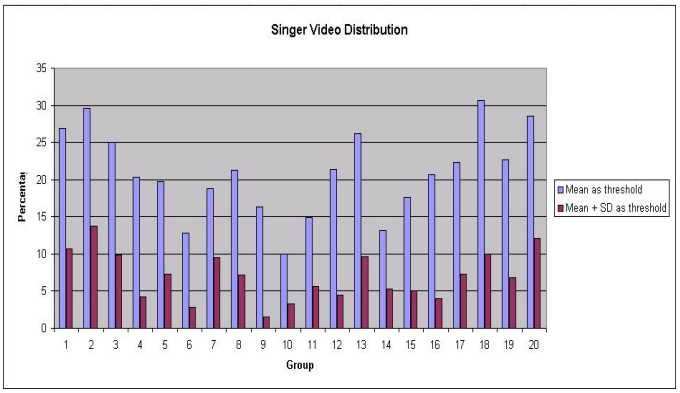

Figure 7. Singer video distribution in YouTube.

		Concept Words from Group			
		A	B	C	D
Percentage of Videos from Group	A	34.04%	4.02%	0%	2.13%
	B	0%	15.79%	0%	0%
	C	0.84%	0.84%	7.58%	0.84%
	D	0%	10.17%	3.39%	5.26%

Table 5. Percentage of videos with tags covering concept words across groups

Group	Concept Words	Precision	
		Mean as Threshold	Mean + SD as Threshold
A	sex photos scandal	21.64%	81.58%
B	old classic memories	61.04%	78.16%
C	sad legend died	35.86%	60.34%
D	together couple two	64.44%	79.82%
Average		45.75%	74.96%

Table 3. Precision of objective clusters

concept of the groups can be told. With this table and table 1, we can easily compare the effectiveness of comment and meta data for concept finding. The tables show that tags, titles or short descriptions are not sufficient for concept discovery of online videos.

Figure 7 illustrates the distribution of Hong Kong singer videos in YouTube according to the proposed algorithm using mean and mean + sd as thresholds. From the figure, we can see that the distribution of videos diverse over different threshold values. With the mean coefficient as the threshold, compared to the video groups resulted from the algorithm with mean + SD coefficient as threshold, larger groups of videos can be obtained. In the other words, algorithm associated with a smaller group assigning threshold would result heavier overlapped video groups.

The video clustering results are evaluated by human experts. To make the evaluation less controvertible, we only show the precisions of objective video groups in Table 3 where groups A, B, C, D are cluster 4, 13, 18, 19 respectively in our clustering. In the table, we noticed that assigning videos to groups with a smaller threshold may sometimes lower the precision. This will be caused in the groups which are very distinct to others. As a video group is too specific, the video-group coefficients to the group hold the extreme values. Also, closely related videos to the distinct group is always much fewer than videos which do not. Hence, videos are condense at the lower extreme side regarding the coefficients distribution. As a result, lowered the mean coefficient and caused the poor precision. On the other hand, we can see that the algorithm which assigns videos into groups with a larger threshold yields far better precisions. The average precision of the larger-threshold clustering groups in the table is 74.96% whilst that of the lower-threshold clustering is just 45.75%. The difference between the precisions resulted from clustering with the two different thresholds reflects the degree of extraordinary of the video group. The larger the difference, the more the special the group is. For example, in group 4, the two precisions differ from each other by a large percentage at about 60%, and from the concept words we can know that this group is about scandal of singers involving their sex photos. This is obviously an extremely distinct group.

7.3 User Comments vs User Tags

As tags are believed to be an accurate description of an object and have been widely used for finding user interests and grouping objects, it is necessary to examine the virtues of user comments over tags before utilizing comments to capture public attentions and categorize videos to facilitate the video search in video sharing sites. One important ob-

Cluster I	Top 10 concept words in user comments	old classic memories drama childhood love 80s memory loved san
	Top 10 frequent user tags	chinese chan mv cheung wong love music mtv top anita
Cluster II	Top 10 concept words in user comments	sad legend two died missed heaven star superstar crying talented
	Top 10 frequent user tags	cheung chan leslie anita mui chinese mv danny hong wong
Cluster III	Top 10 concept words in user comments	guitar solo band rock cover drummer chords intro crap violin
	Top 10 frequent user tags	chinese beyond wong kong cheung ka kui hong nicholas paul
Cluster IV	Top 10 concept words in user comments	sex photos stupid fake victims private innocent scandal girls stop
	Top 10 frequent user tags	gillian chung sex photo edison chen gill cheung cecilia chan

Table 4. Examples of concept words from user comments and user tags in four video clusters

servation from our experimental results is that user comments usually contains more in-depth information than that of user tags. Table 4 shows both the top 10 concept words found from user comments and the top 10 user tags of four clustered groups. From the concept words in the user comments, we can make a reasonable prediction that cluster I is about some music videos of some old songs. From the user tags, however, we can only find some singer names or some high-level descriptions (e.g. music, mv, mtv). Same as cluster II, from the concept words, this cluster is probably talking about some superstars who are already died. Nevertheless, the most frequent tags are only names of those dead superstars which do not reveal the low-level description of the group. Cluster III is the similar case as the above two clusters. Concept words from user comments state that this group is about the band sound and rock music but the tags only list out the name of a local popular band, "Beyond", and some of the band members. Tags of the other clusters suffer the similar problem as the above mentioned clusters. From the table, we can see that the user tags actually agree with our discovered concept words though the tags just exhibit the high-level sketch of the groups. In the other words, our algorithm gives an in-depth characterization of the videos with the concept words which the characterization cannot be exposed by the user tags, and in the mean time, the concept words achieve a strong agreement with the tags.

From this observation, we can conclude that if we want to obtain clustering results in a more fine-grained level, using commentary-based clustering technique is more suitable. For the purpose of facilitating video search, it is beyond doubt that result of fine-grained level clustering involving user points of attention is more desirable.

To give a more in-depth analysis of comments and tags, we have compared concept words against tags in different clusters. Table 5 records the portion of videos whose tags cover the concept words of different groups and there are two major observations from the table. First, we can see that there are at least 65% of videos whose tags cannot cover the concept words of the group they belongs to. This implies tag-based clustering cannot completely capture user opinions and video content. Second, we can see that the concept words of each group are mostly covered by tags of its own group. This once again verify the accuracy of our proposed method.

8 Conclusion and Future Work

In this paper, we have proposed a novel commentary-based matrix factorization technique to cluster videos to facilitate searching and generate concept words to improve indexing. We propose the use of factorized component entropy as a measure to construct good vocabulary for analyzing sparse and noisy social media data. Experimental results showed that our commentary-based clustering yields better performance than that of tag-based approach which was proposed previously in the literature. On the other hand, we have successfully discovered some non-trivial categories among the videos of Hong Kong singers. Since our categorization is learnt from user feedbacks, it can provide an easy way for users to reach their desired videos via our list of categories.

In our future work, we plan to extend the commentary-based technique from video clustering to user and singer clustering. After we have obtained the three types of clusters, we can acquire the relationships among different videos, singers and users by analyzing the inter-cluster similarity. As such, social culture can be studied by combining and analyzing the discovered relationships. With the video-video, singer-singer, user-singer, and user-user relationships found by clustering, we can know the changes in

music styles and singer styles over the ages, the trend of music, the ways people appreciate music, and even the special relationships of singers reflected by news, and more. Relationships observed by clustering are not only useful for social scientists to study social culture, but also beneficial for businesses, entertainment companies, fans clubs, social network systems and system users. With the help of examined user-user relationships, businesses can be profited from reducing advertising costs by advertise only to the potential customer groups. User-signer relationships define user-idol groups, entertainment companies can effectively promote to the target groups. Determining the user-singer relationships, in addition to profits for entertainment companies, fans groups can easily be managed. Other than the advantages for some specific parties, general users are also benefited. Well-clustered groups of videos and singers equipped with a batch of concept words leads to a effort saving video searching for users. Also, social network systems are able to detect and refine incorrect tags with the concept words resulted from clustering. As a result, description of videos are more precise and thus improves the video searching function.

References

[1] http://www.bebo.com.

[2] http://www.dailymotion.com.

[3] http://www.facebook.com.

[4] http://www.flickr.com.

[5] http://www.myspace.com.

[6] http://www.youtube.com.

[7] Usa today. youtube serves up 100 million videos a day online.

[8] F. Benevenuto, F. Duarte, T. Rodrigues, V. A. Almeida, J. M. Almeida, and K. W. Ross. Understanding video interactions in youtube. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, pages 761–764, New York, NY, USA, 2008. ACM.

[9] X. Cheng, C. Dale, and J. Liu. Understanding the characteristics of internet short video sharing: Youtube as a case study. *In CoRR abs*, Jul 2007.

[10] G. Geisler and S. Burns. Tagging video: conventions and strategies of the youtube community. In *JCDL '07: Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 480–480, New York, NY, USA, 2007. ACM.

[11] L. Guo, S. Jiang, L. Xiao, and X. Zhang. Fast and low-cost search schemes by exploiting localities in p2p networks. *J. Parallel Distrib. Comput.*, 65(6):729–742, 2005.

[12] M. J. Halvey and M. T. Keane. Exploring social dynamics in online media sharing. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 1273–1274, New York, NY, USA, 2007. ACM.

[13] M. Heckner, T. Neubauer, and C. Wolff. Tree, funny, to read, google: what are tags supposed to achieve? a comparative analysis of user keywords for different digital resource types. In *SSM '08: Proceeding of the 2008 ACM workshop on Search in social media*, pages 3–10, New York, NY, USA, 2008. ACM.

[14] J. Heer and D. Boyd. Vizster: Visualizing online social networks. *IEEE Symposium on Information Visualization, 2005*, 2005.

[15] C. M. C. Y. Julia Stoyanovich, Sihem Amer-Yahia. Leveraging tagging to model user interests in del.icio.us. In *AAAI '08: Proceedings of the 2008 AAAI Social Information Spring Symposium*. AAAI, 2008.

[16] P. P. Kotsiantis S., Kanellopoulos D. Multimedia mining. In *WSEAS Transactions on Systems, Issue 10, Volume 3*, pages 3263–3268, December 2004.

[17] X. Li, L. Guo, and Y. E. Zhao. Tag-based social interest discovery. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 675–684, New York, NY, USA, 2008. ACM.

[18] J. Z. Pan, S. Taylor, and E. Thomas. Reducing ambiguity in tagging systems with folksonomy search expansion. In *ESWC 2009 Heraklion: Proceedings of the 6th European Semantic Web Conference on The Semantic Web*, pages 669–683, Berlin, Heidelberg, 2009. Springer-Verlag.

[19] C. G. R. A. A. F. L. Rodrygo L. T. Santos, Bruno P. S. Rocha. Characterizing the youtube video-sharing community. 2007.

[20] M. F. Schwartz and D. C. M. Wood. Discovering shared interests using graph analysis. *Commun. ACM*, 36(8):78–89, 1993.

[21] A. S. Sharma and M. Elidrisi. Classification of multimedia content (video's on youtube) using tags and focal points. Working paper.

[22] K. Sripanidkulchai, B. Maggs, and H. Zhang. Efficient content location using interest-based locality in peer-to-peer systems. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE*, volume 3, pages 2166–2176 vol.3, 2003.

[23] S. Tsekeridou and I. Pitas. Content-based video parsing and indexing based on audio-visual interaction, 2001.

[24] W. Xu, X. Liu, and Y. Gong. Document clustering based on non-negative matrix factorization. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 267–273, New York, NY, USA, 2003. ACM.

[25] L. Yang, J. Liu, X. Yang, and X.-S. Hua. Multi-modality web video categorization. In *MIR '07: Proceedings of the international workshop on Workshop on multimedia information retrieval*, pages 265–274, New York, NY, USA, 2007. ACM.

[26] O. R. Zaïane, J. Han, Z.-N. Li, S. H. Chee, and J. Y. Chiang. Multimediaminer: a system prototype for multimedia data mining. In *SIGMOD '98: Proceedings of the 1998 ACM SIGMOD international conference on Management of data*, pages 581–583, New York, NY, USA, 1998. ACM.