

# Adaptive Association Rule Mining for Web Video Event Classification

Chengde Zhang<sup>1</sup>, Xiao Wu<sup>1</sup>, Mei-Ling Shyu<sup>2</sup>, Qiang Peng<sup>1</sup>

<sup>1</sup>School of Information Science and Technology  
Southwest Jiaotong University, Chengdu, 610031, P.R.China

<sup>2</sup>Department of Electrical and Computer Engineering  
University of Miami, Coral Gables, FL 33146, USA  
chengde66@gmail.com, {wuxiaohk,qpeng}@home.swjtu.edu.cn, shyu@miami.edu

## Abstract

*Due to the popularity and development of social networks and web video sites, we have witnessed an exponential growth in the volumes of web videos in the last decade. This prompts an urgent demand for efficiently grasping the major events. Nevertheless, the insufficient and noisy text information has made it difficult and challenging to mine the events based on the initial keywords and visual features. In this paper, we propose an adaptive semantic association rule mining method in the NDK (Near-Duplicate Keyframes) level to enrich the keyword information and to remove the words without any semantic relationship. Moreover, both textual and visual information are employed for event classification, targeting for bridging the gap between NDKs and the high-level semantic concepts. Experimental results on large scale web videos from YouTube demonstrate that our proposed method achieves good performance and outperforms the selected baseline methods.*

**Keywords:** Web Video Event Classification, Adaptive Association Rule Mining, Near-Duplicate Keyframes (NDK).

## 1. Introduction

Due to the fast advancement of multimedia technologies, the number of web videos has been growing explosively in the past decade. In China, there are many video-distributing web video sites emerged in the past several years like YouKu, TuDou, etc. It is easier for the users to get a large number of relevant web videos about ongoing incidents or events through search engines or video sharing websites.

Generally, many users prefer to watch web videos, rather than from the news documents, to get the most updated status or progress of major events and hot topics. Moreover, most of the users first want to know the major events and

then to construct their relationship in minds. However, it is still a challenge for the users to grasp the major events after a glance at the search results. In fact, it is really daunting for the users to view from one video to another one to select the hot topics (events of interests), not to mention for them to manually summarize the videos after watching all of them at the same time. This is not only highly time consuming but also difficult for them, especially for the unfamiliar topics. Therefore, it is becoming critical to automatically group relevant web videos together.

To mine events, both visual and textual information contain rich information. Generally speaking, for visual information, important shots might be frequently inserted into videos as a reminder or support of viewpoints, while useful visual information is usually embedded in them. According to visual content information, these duplicate keyframes can be classified to groups, where each group is called an NDK (near-duplicate keyframe)[10] group. NDKs are useful information to mine videos with similar contents. However, for the textual information, web videos generally contain much less text information than traditional documents, because the number of user-supplied text information (title and tags) is limited. Moreover, even with titles and tags, they still exhibit noisy, ambiguous, and sometimes incomplete features. To make matters worse, the users might add irrelevant hot terms (words) to attract the attention. As a result, the discriminative capability of text characteristics is exacerbated in the web scenes. This motivates us to explore the grouping of semantic related terms together through an adaptive association rule mining (ARM) method. It is believed that making use of the grouped terms to find the statistics and distribution characteristics in near-duplicate keyframe group can improve the performance of event classification to enhance their robustness.

In this paper, we propose a novel framework, which discovers the statistics and the distribution characteristics of

the grouped terms in clustered keyframes for web video event classification. In our proposed framework, an adaptive association rule mining method is first developed to address the issues related to the noisy data and the scarce textual information. Second, transitive closure is applied to group the terms together, while the semantic relationships between terms help better bridge the gap between NDKs and the high-level semantic concepts. Finally, three classifiers are used to evaluate the NDK-level event similarity with the assistance of the textual information that are more stable than the single terms. Moreover, both terms and NDKs with relatively low frequencies are still considered as useful information in the experiments, while the web videos without duplicate content information are taken as noise.

The rest of this paper is organized as follows. Section 2 gives a brief overview of the related work. The details of the framework for web video event classification is discussed in Sections 3. The experiments and results are presented in Section 4. Finally, Section 5 concludes this paper.

## 2. Related Work

### 2.1 Topic Detection and Tracking

The goal of topic detection and tracking (TDT) is to detect new topics and track the known events in text news streams, while an event is something that happens at a specific time and place, along with all the necessary preconditions and unavoidable consequences [3]. In fact, web video event classification is considered to belong to the task of TDT. In the text research areas, much work [5, 6, 8, 25, 32] has been done about TDT. There are also many research studies about TDT in the multimedia research areas. For example, topics are tracked with visual duplicates and semantic concepts [9, 16]. The concept of near-duplicate keyframes (NDK) has abundantly used in real applications [23, 31]. An NDK is a set of similar keyframes but with certain variations induced by the acquisition times, lighting conditions, and editing operations [10]. NDK detection is to calculate the keyframe similarity among videos. The retrieval of NDKs plays an important role in measuring video clip similarity and tracking video shots of multi-lingual sources [10, 20, 29, 35]. With the assistance of NDK constraints, news stories are clustered into topics by constraint-based co-clustering [30].

Recently, a very popular trend is to integrate both textual and visual information for web video event classification. For example, a trajectory-based approach presented in [2] is used to discover, track, monitor, and visualize web video topics. With textual correlation and keyframe matching, topic clusters are grouped in [4] and news stories from different TV channels are linked in [34]. Topic discovery is deployed by constructing the duality between stories and

textual-visual concepts through a bipartite graph [17, 29]. Events are discovered by text co-occurrence and visual feature trajectory in [28].

### 2.2 Association Rule Mining (ARM)

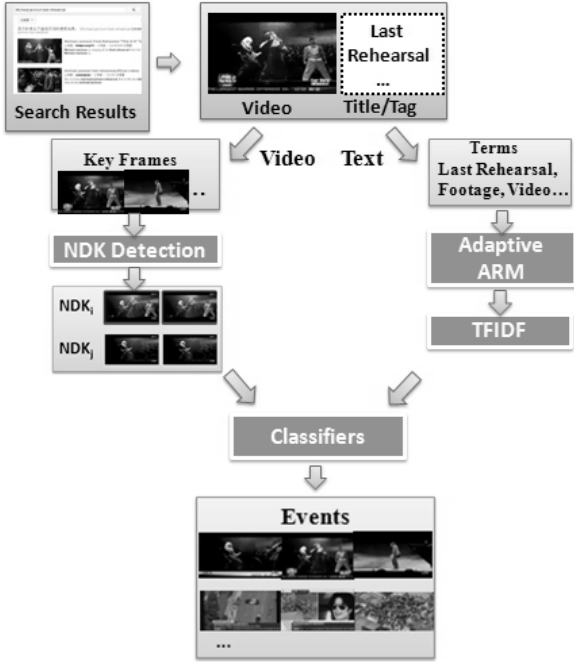
Association rule mining (ARM) is known to discover interesting and useful co-occurring associations among data in large databases [1, 22], and has also been widely adopted in multimedia research to bridge the semantic gap between low-level multi-modal features and the concepts of interest. The number of discovered association rules in a data mining application is typically huge, which makes it difficult for the users to identify those rules that are of particular interest to them. This calls for the need to remove the insignificant rules so that more interesting rules can be utilized.

There are three categorized interestingness measures for generating and selecting rules: objective measures, subjective measures, and semantic measures as described in [7]. In [13], the authors mentioned that the objective measures can be calculated based on probability, statistics, distance, or information theory, while most of the criteria depend only on the data characteristics. The objective measures can be used to prune the rules for web image clustering, as instructed in [19]. On the other hand, an approach is introduced in [11] to aggregate a set of objective interestingness measures using the Choquet integral as the aggregation operator to find the most interesting association rules. A interestingness analysis system (IAS) is developed to assist the users in finding unexpected rules from a set of discovered association rules [15]. A ShotRank notion is introduced in [33] as a measure of subjective interestingness measure for a video browsing and summarization system. The authors in [12] introduced the utilization of Multiple Correspondence Analysis (MCA) as a utility-based semantic measure for association rule generation.

Though ARM has been utilized in many applications, it has not been utilized in finding the associations of the visual features of NDKs. This motivates the study of using the distribution characteristics information of terms in NDKs to improve the performance of event classification, especially an adaptive association rule mining approach.

## 3. Proposed Framework

This paper proposes a novel framework that performs web video classification via the use of an adaptive association rule mining (ARM) method for semantic relationship detection. The proposed framework is illustrated in Figure 1. As can be seen from this figure, it consists of three stages, namely data preprocessing, adaptive association rule mining, and classification.



**Figure 1. The proposed web video event classification framework.**

In the data preprocessing stage, near-duplicate keyframes (NDKs) are first detected from keyframes of the web videos. For the visual data sets, to ensure good performance of NDK detection between videos, local points are detected with Harris-Laplace and are described by SIFT [18]. The public available tool proposed in [36] is adopted to detect the NDKs. The detected NDKs are further grouped to form clusters by transitive closures. As NDKs have the unique property of identifying similar events, all NDKs are considered as the features. For the textual data sets, the terms extracted from titles and tags are used as textual features. Due to the noisy user-supplied tag information, special characters (e.g., #, >, ?, |) are first removed, and then text words are pruned by word stemming, Chinese word segmentation, and so on. Finally, all features and the class labels are combined to form an indicator matrix with instances (i.e., NDKs) as rows and the categories of variables (i.e., Terms) and the concepts (or class labels) as columns.

In the adaptive association rule mining stage, to better discover the association patterns from the noisy terms, an adaptive association rule mining method is first developed, where an adaptive support value is selected for different lengths of items according to the F1-score value calculated by the classifiers. Next, transitive closure is used to combine the related terms to groups.

In the classification stage, every NDK is grouped into

the event with the largest similarity. In order to prove the effectiveness of the proposed method, the K-Nearest Neighbor (KNN), NaiveBayes (NB), and Support Vector Machine (SVM) classifiers are applied to calculate the similarity between each NDK and the class, respectively.

### 3.1. Adaptive Association Rule Mining

Hot terms are frequently appeared in related videos. Since semantic related terms usually carry useful information, they can be used to group related videos, even with similar themes but different visual content. However, due to multi-language, multi-semantics, polysemy and synonyms problems, highly related web videos might be annotated with different words. There are numerous frequently accompanied terms that convey useful information, and thus it makes sense that semantic related terms can be used to identify related web videos together. For this purpose, ARM is applied to improve the performance by pruning the noise and strengthen the important terms and their semantically related words with higher weights.

$$T = \langle T_1, T_2, T_3, \dots, T_n \rangle \quad (1)$$

Let  $T$  be a set of terms (as shown in Equation (1)). We first use the traditional ARM to get the candidates of k-itemsets as shown in Table 1. Next, let  $T_i$  and  $T_j$  be two Terms. The rationale is that the more common videos they contain, the more similar they are. Due to the different frequencies of the terms, taking the frequency as the support value might lose a large number of Non-Hot Words which may contain useful information. Therefore, to capture the support value, the Jaccard coefficient is adopted to measure the similarity between  $T_i$  and  $T_j$ , as shown in Equation (2).

$$Support(T_i, T_j) = \frac{|T_i \cap T_j|}{Min(T_i, T_j)} \quad (2)$$

where  $|T_i \cap T_j|$  is the number of overlapped videos, and  $Min(T_i, T_j)$  is the minimum number of videos contained in  $T_i$  and  $T_j$ .

Since there are different frequencies in the term pairs, using the same threshold value may be improper. In order to find the best support value for different lengths of items, the support value distribution range is first divided into 10 segments (0.1 to 1.0) as shown in Figure 2, with the horizontal and vertical axes representing the support and F1-score values, respectively. Then, the support values are sorted in an ascending order without the duplicate values for the training data set. In each iteration, a different threshold value is applied to the training data set. Finally, the one with the highest F1-score (evaluated by the classifiers) is selected as the support value. For example, in Figure 2, the value of the 9th (0.9) interval is chosen as the threshold value as it achieves the highest F1 score.

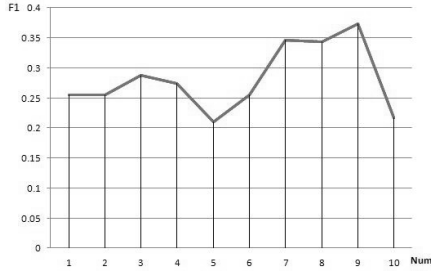


Figure 2. Adaptive support value selection.

Table 1. Frequent itemsets from adaptive ARM

1-Items	2-Items	...	N-Items
$T_1$	$T_1, T_2$	...	$T_1, T_2, T_3 \dots$
$T_2$	$T_2, T_3$	...	$T_5, T_6, T_7 \dots$
$T_3$	$T_5, T_6$	...	$T_{11}, T_{12}, T_{13} \dots$
...	...	...	...

Common terms between two term groups ( $G_i$  and  $G_j$ ) denote that they are still high correlated. If the score is above a certain threshold value, they can be regarded as similar, denoted as  $G_i \rightarrow G_j$ . This is a symmetric measure, which also implies that  $G_j \rightarrow G_i$ . For certain cases, there may also be relationships among term groups even if they have no common videos. This is because that they represent the same theme of the videos. Hence, if  $G_i \rightarrow G_j$  and  $G_j \rightarrow G_k$ , then we assume that there exists the association  $G_i \rightarrow G_k$ . Therefore, the transitive closure is used to combine related term groups together.

### 3.2. Classification

After obtaining the grouped terms ( $G_i$ ), the correlation between grouped terms  $G_i$  and near-duplicate keyframe groups  $NDK_j$  can be expressed in the form of matrix as shown in Table 2, which is represented in a 2-dimensional NDK-Term matrix (NT), where each element  $NT_j^i$  in  $NT$  is defined in Equation (3).

$$NT_j^i = \frac{F(i, j)}{\sum N(j)} \times \log \frac{N}{D(G_i)}, \quad (3)$$

where  $F(i, j)$  is the frequency of term group  $G_i$  appeared in  $NDK_j$ ,  $\sum N(j)$  is the frequency of all the term groups appeared in  $NDK_j$ ,  $D(G_i)$  is the number of NDKs that contain  $G_i$ , and  $N$  is the total number of NDKs.

Next, data instances are split into a training set and a testing set. For each class, one half of the data set are randomly selected as the training set, while the other half are taken as the testing set. The main reasons of this process can be shown as follows. On one hand, the splitting is able to mitigate any bias caused by the particularly chosen instances;

Table 2. Training dataset after terms grouping

	$G_1$	...	$G_N$	Class
$NDK_1$	$NT_1^1$	...	$NT_1^N$	0
$NDK_2$	$NT_2^1$	...	$NT_2^N$	1
...	...	...	...	...

On the other hand, each class being properly represented in both training and testing sets could be guaranteed.

Finally, in order to prove our proposed adaptive ARM is an effective method, three popular classifiers are used. WEKA [26, 27] provides the implementations of these three classifiers, namely K-Nearest Neighbor (KNN) where  $K=3$ , Native Bayes (NB), and Support Vector Machine (Sequential Minimal Optimization) (SVM).

## 4. Experiments and Results

### 4.1. Data Set Information

We use the data set given by [28] for evaluation, which consists of 19,972 web videos and 803,346 keyframes. These videos are from YouTube and most of the videos were collected in May 2009. The topics are selected based on the top 10 news happened during 2006 to 2009 as recommended by CNN, TIMES, and Xinhua. We select 14 hot topics shown in Table 3, which basically cover different characteristics of topics for the experimental evaluation. For example, topic "Virginia tech massacre" only spans for one month from April to May of year 2007. However, the topic "California wildfires" happened periodically for several years, when the content is relatively homogeneous. To ensure fairness, those events having fewer than five web videos are regarded as noise and pruned out. Therefore, 10,716 videos, 35,555 NDKs, and 41,814 terms are used as the data set in the experiments. The detailed information of the dataset is listed in Table 3. Each topic is composed of several events. Take topic "Russian President Election" as an example, where it contains 6 events: "election related news", "Putin and Medvedev news", "Election Day", "Comment After Election", "Medvedev sworn in news", and "Medvedev president video". For the ground truth, we collected the data by searching each topic from Wikipedia and Google, and then manually decided the ground truth.

### 4.2. Evaluation of The Framework

The Precision (P), Recall (R), and F1 measure (F1) are selected to evaluate the performance of event classification, as defined in Equations (4), (5), and (6).

$$Precision = \frac{|B_i^+|}{|A_i|} \quad (4)$$

**Table 3. Dataset information**

ID	Topic	Videos#	NDK#	Terms#	Events#
1	Beijing Olympics	1,098	4,593	4,861	17
2	Mumbai terror attack	423	1,741	1,569	5
3	Russia Georgia war	749	2,823	2,316	7
4	Somali pirates	410	1,405	2,178	5
5	Virginia tech massacre	683	1,865	1,621	2
6	Israel attacks Gaza	802	3,087	3,546	4
7	Beijing Olympic torch relay	652	2,448	1,949	12
8	Sichuan earthquake	1,458	4,963	4,806	6
9	California wildfires	426	1,631	3,025	6
10	London terrorist attack	784	2,183	4,228	5
11	Oil price	759	2,486	3,814	5
12	Myanmar cyclone	613	2,698	1,624	4
13	Kosovo independence	524	969	1,593	5
14	Russian president election	1,335	3,930	4,684	6
Total		10,716	35,555	41,814	89

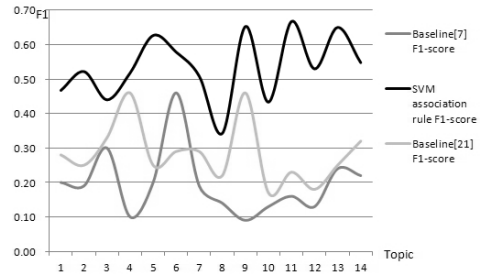
$$Recall = \frac{|B_i^+|}{|B_i|} \tag{5}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

where  $B_i^+$  is the number of correctly grouped positive videos for cluster  $A_i$ , and  $B_i$  is the number of positive samples in the ground truth.

To fully evaluate our framework, we first evaluate the performance of the proposed adaptive ARM method with three classifiers(KNN, SVM, NB). In order to be fair to the other classifiers, the original features and default parameter setting in WEKA are used. Next, the comparison is conducted among the results of the adaptive ARM method against two baseline methods, namely the simplified versions of the proposed methods in [8] and [28]. The performance comparison of the F1-score values is presented in Figure 3, with the abscissa and ordinate representing the topics and F1-score values, respectively. Due to the trade-off between Precision and Recall, a better F1-score value is usually considered as the better performance to demonstrate the accuracy of a framework. Hence, the focus of the performance comparison is on the F1 measure.

Tables 4, 5, and 6 show the Precision, Recall and F1-score values of fourteen topics for each classifier by using the original single term, term clustering method in [32], and our proposed framework with the adaptive ARM method, respectively. For each topic, the best Precision, Recall and F1 are highlighted in bold in these tables. As can be seen from these three tables that the average precision values of our proposed framework are higher than the other two methods in most of the instances. However, the single term approach gets a better average Recall value, since it is inevitable that there are missing words with useful information. Moreover, our framework obtains better F1-score values than the other two methods, with the average maximum



**Figure 3. Comparison of F1-scores.**

value up to 54% in Table 5.

The method of text feature trajectories in [8], demonstrates that text information alone is not enough for event classification. In [28], its performance is better than the method in [8]. However, it can be seen that its performance is still not good enough. Because it misses those low-frequency terms and NDKs, while they contain a large number of web videos information. The visual feature trajectory of NDKs is not consistent. Hence, the authors in [28] attempt to use term co-occurrence to compensate the defects for visual information. However, the frequent patterns miss some information for event classification too.

Figure 3 shows the F1-score performance comparison. It is clear to see that our proposed framework achieves promising results compared to both baseline methods since the F1 values of our proposed framework are significantly higher. It is even more encouraging to see that the best result reaches 67%, with similar time and space complexities. This can be inferred from the results that an important event is often accompanied with a set of representative terms. Hence, semantic related terms are good cues for grouping related videos into events. The proposed method can be easily extended to a huge data set, though it may in-

**Table 4. Performance comparison - KNN**

Topic	<i>TFIDF (KNN)</i>			<i>CC (KNN) [32]</i>			<i>Association Rule (KNN)</i>		
	P	R	F1	P	R	F1	P	R	F1
1	0.41	<b>0.44</b>	0.42	0.48	0.39	0.43	<b>0.65</b>	0.37	<b>0.47</b>
2	0.31	<b>0.52</b>	0.39	0.34	0.50	0.41	<b>0.42</b>	0.50	<b>0.46</b>
3	0.55	0.24	0.34	<b>0.64</b>	0.25	0.36	0.55	<b>0.38</b>	<b>0.45</b>
4	0.32	0.54	0.40	<b>0.41</b>	<b>0.60</b>	<b>0.49</b>	0.40	0.58	0.48
5	0.69	0.80	0.74	<b>0.71</b>	<b>0.79</b>	<b>0.75</b>	<b>0.71</b>	0.78	0.74
6	0.40	0.72	0.52	0.44	0.61	0.51	<b>0.48</b>	<b>0.62</b>	<b>0.54</b>
7	0.34	0.53	0.42	0.48	<b>0.51</b>	<b>0.49</b>	<b>0.57</b>	0.25	0.35
8	0.55	<b>0.40</b>	0.47	0.46	0.35	0.40	<b>0.64</b>	0.39	<b>0.49</b>
9	0.49	0.45	0.47	<b>0.67</b>	0.31	0.42	0.56	<b>0.71</b>	<b>0.63</b>
10	0.21	<b>0.75</b>	0.33	<b>0.33</b>	0.67	<b>0.45</b>	0.32	0.67	0.43
11	0.23	<b>0.78</b>	0.35	0.45	0.46	0.46	<b>0.62</b>	0.55	<b>0.58</b>
12	0.33	<b>0.62</b>	0.43	<b>0.39</b>	0.57	0.46	<b>0.39</b>	0.59	<b>0.47</b>
13	<b>0.84</b>	<b>0.50</b>	<b>0.62</b>	0.75	0.38	0.50	0.80	0.48	0.60
14	0.30	0.50	0.38	0.32	<b>0.53</b>	0.40	<b>0.83</b>	0.33	<b>0.47</b>
<b>Average</b>	0.43	<b>0.56</b>	0.45	0.49	0.49	0.47	<b>0.57</b>	0.51	<b>0.51</b>

**Table 5. Performance comparison - SVM**

Topic	<i>TFIDF (SVM)</i>			<i>CC (SVM) [32]</i>			<i>Association Rule (SVM)</i>		
	P	R	F1	P	R	F1	P	R	F1
1	0.53	<b>0.44</b>	<b>0.48</b>	0.62	0.38	0.47	<b>0.68</b>	0.36	0.47
2	0.31	0.46	0.37	0.42	0.49	0.45	<b>0.44</b>	<b>0.63</b>	<b>0.52</b>
3	0.49	0.17	0.25	<b>0.68</b>	0.30	0.41	0.60	<b>0.35</b>	<b>0.44</b>
4	0.34	<b>0.59</b>	0.43	0.40	0.55	0.46	<b>0.46</b>	<b>0.59</b>	<b>0.52</b>
5	0.69	<b>0.80</b>	0.74	0.71	<b>0.79</b>	<b>0.75</b>	0.65	0.61	0.63
6	0.49	<b>0.67</b>	0.56	0.47	0.64	0.54	<b>0.54</b>	0.62	<b>0.58</b>
7	0.47	0.48	0.47	0.59	<b>0.53</b>	<b>0.56</b>	<b>0.65</b>	0.42	0.51
8	0.51	0.30	0.38	<b>0.69</b>	<b>0.49</b>	<b>0.58</b>	0.46	0.27	0.34
9	0.50	<b>0.55</b>	0.53	0.70	0.32	0.44	<b>0.88</b>	0.52	<b>0.65</b>
10	0.22	0.68	0.33	<b>0.34</b>	<b>0.74</b>	<b>0.46</b>	0.32	0.69	0.43
11	0.30	0.63	0.41	0.56	0.59	0.57	<b>0.63</b>	<b>0.71</b>	<b>0.67</b>
12	0.31	0.60	0.41	0.40	0.66	0.50	<b>0.47</b>	<b>0.61</b>	<b>0.53</b>
13	0.83	0.49	0.61	0.82	0.42	0.55	<b>0.90</b>	<b>0.51</b>	<b>0.65</b>
14	0.34	0.63	0.44	0.43	0.50	0.46	<b>0.48</b>	<b>0.64</b>	<b>0.55</b>
<b>Average</b>	0.45	<b>0.53</b>	0.46	0.56	0.53	0.51	<b>0.58</b>	<b>0.53</b>	<b>0.54</b>

crease the time complexity in the training phase for more textual information processing. However, it can reduce the space complexity as the dimension of the textual information is smaller. Compared to the traditional Adaptive ARM in [14, 21, 24], on one hand, we take F1-score values to evaluate the results to consider both the Precision and Recall values; on the other hand, the method of calculating the support value takes into account the characteristics of textual information about web videos. That is, high-frequency and low-frequency terms may also have a high relationship due to different writing habits of web users.

## 5. Conclusion

With the fast development of the Internet, it becomes an pressing need for the users to be able to quickly find out the major event. Considering the distinct characteristics of the web video, namely the limited features, the unavoidable NDK detection problem, and the noisy and less text information, web video event classification has always been a challenging task. In this paper, a novel adaptive ARM-based framework is proposed for web video event classification, which applies the proposed adaptive ARM method to address the noisy and less textual information issues. Next, semantic textual information is used to help better bridge the gap between NDKs and the high-level semantic concepts. Moreover, both textual and visual features with relatively low frequencies are considered useful information in our proposed framework.

## Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (No. 61071184, No. 60972111, No. 61036008), Research Funds for the Doctoral Program of Higher Education of China (No. 20100184120009), Program for Sichuan Provincial Science Fund for Distinguished Young Scholars (No. 2012JQ0029, No. 13QNJJ0149), the Fundamental Research Funds for the Central Universities (Project No. SWJTU09CX032, No. SWJTU10CX08), and financial support from the program of China Scholarships Council (No. 201207000050). Chengde Zhang gratefully acknowledges the opportunity of the exchange visiting scholar program from the University of Miami. Special thanks go to Dianting Liu and Qiusha Zhu for their great help in this study.

## References

[1] R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 207–216, 1993.

[2] J. Cao, C.-W. Ngo, Y.-D. Zhang, and J.-T. Li. Tracking web video topics: Discovery, visualization, and monitoring. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(12):1835–1846, 2011.

[3] K.-Y. Chen, L. Luesukprasert, S. Chou, et al. Hot topic extraction based on timeline analysis and multidimensional sentence modeling. *IEEE Transactions on Knowledge and Data Engineering*, 19(8):1016–1025, 2007.

[4] P. Duygulu, J.-Y. Pan, and D. A. Forsyth. Towards auto-documentary: Tracking the evolution of news stories. In *Proceedings of the 12th ACM International Conference on Multimedia*, pages 820–827, 2004.

[5] G. P. C. Fung, J. X. Yu, H. Liu, and P. S. Yu. Time-dependent event hierarchy construction. In *Proceedings of the 13th ACM International Conference on Knowledge Discovery and Data Mining*, pages 300–309, 2007.

[6] G. P. C. Fung, J. X. Yu, P. S. Yu, and H. Lu. Parameter free bursty events detection in text streams. In *Proceedings of the 31st International Conference on Very Large Data bases*, pages 181–192, 2005.

[7] L. Geng and H. J. Hamilton. Interestingness measures for data mining: A survey. *ACM Computing Surveys*, 38(3):9, 2006.

[8] Q. He, K. Chang, and E.-P. Lim. Analyzing feature trajectories for event detection. In *Proceedings of the 30th ACM International Conference on Research and Development in Information Retrieval*, pages 207–214, 2007.

[9] W. H. Hsu and S.-F. Chang. Topic tracking across broadcast news videos with visual duplicates and semantic concepts. In *Proceedings of the 2006 IEEE International Conference on Image Processing*, pages 141–144, 2006.

[10] Y. Ke, R. Sukthankar, and L. Huston. Efficient near-duplicate detection and sub-image retrieval. In *ACM Multimedia*, volume 4, page 5, 2004.

[11] T. T. N. Le, H. X. Huynh, and F. Guillet. Finding the most interesting association rules by aggregating objective interestingness measures. In *Knowledge Acquisition: Approaches, Algorithms and Applications*, pages 40–49. Springer, 2009.

[12] L. Lin, G. Ravitz, M.-L. Shyu, and S.-C. Chen. Correlation-based video semantic concept detection using multiple correspondence analysis. In *Proceedings of the Tenth IEEE International Symposium on Multimedia*, pages 316–321, 2008.

[13] L. Lin, M.-L. Shyu, and S.-C. Chen. Association rule mining with a correlation-based interestingness measure for video semantic concept detection. *International Journal of Information and Decision Sciences*, 4(2):199–216, 2012.

[14] W. Lin, S. A. Alvarez, and C. Ruiz. Collaborative recommendation via adaptive association rule mining. In *Proceedings of the International Workshop on Web Mining for E-Commerce*, 2000.

[15] B. Liu, W. Hsu, S. Chen, and Y. Ma. Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems and Their Applications*, 15(5):47–55, 2000.

[16] D. Liu, M.-L. Shyu, C. Chen, and S.-C. Chen. Within and between shot information utilisation in video key frame extraction. *Journal of Information & Knowledge Management*, 10(3):247–259, 2011.

**Table 6. Performance comparison - NB**

Topic	TFIDF (NB)			CC (NB) [32]			Association Rule (NB)		
	P	R	F1	P	R	F1	P	R	F1
1	0.47	<b>0.46</b>	<b>0.46</b>	0.51	0.37	0.43	<b>0.65</b>	0.33	0.44
2	0.24	0.49	0.32	0.35	0.38	0.36	<b>0.47</b>	0.46	<b>0.47</b>
3	0.61	0.26	0.36	0.68	0.26	0.37	<b>0.70</b>	<b>0.29</b>	<b>0.41</b>
4	0.34	0.57	0.43	<b>0.37</b>	0.52	0.43	<b>0.37</b>	<b>0.67</b>	<b>0.48</b>
5	0.69	0.80	0.74	<b>0.71</b>	<b>0.79</b>	<b>0.75</b>	<b>0.71</b>	0.77	0.74
6	<b>0.43</b>	<b>0.74</b>	<b>0.55</b>	0.37	0.42	0.39	0.40	0.55	0.46
7	0.37	0.42	0.39	0.45	<b>0.54</b>	<b>0.49</b>	<b>0.46</b>	0.46	0.46
8	0.56	0.43	0.48	0.57	<b>0.46</b>	<b>0.51</b>	<b>0.63</b>	0.41	0.49
9	0.47	<b>0.61</b>	0.53	0.60	0.23	0.33	<b>0.68</b>	0.55	<b>0.61</b>
10	0.20	<b>0.78</b>	0.32	0.31	0.63	0.41	<b>0.34</b>	0.60	<b>0.44</b>
11	0.27	<b>0.85</b>	0.41	0.48	0.25	0.33	<b>0.59</b>	0.58	<b>0.58</b>
12	0.32	0.63	0.43	<b>0.35</b>	0.58	<b>0.44</b>	<b>0.35</b>	<b>0.60</b>	<b>0.44</b>
13	0.77	<b>0.33</b>	<b>0.46</b>	<b>0.89</b>	0.20	0.33	0.81	0.18	0.30
14	0.24	<b>0.41</b>	0.30	0.35	0.34	0.34	<b>0.60</b>	0.36	<b>0.45</b>
<b>Average</b>	0.43	<b>0.56</b>	0.44	0.50	0.43	0.43	<b>0.56</b>	0.49	<b>0.49</b>

- [17] L. Liu, L. Sun, Y. Rui, Y. Shi, and S. Yang. Web video topic discovery and tracking via bipartite graph reinforcement model. In *Proceedings of the 17th ACM International Conference on World Wide Web*, pages 1009–1018, 2008.
- [18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [19] H. H. Malik and J. R. Kender. Clustering web images using association rules, interestingness measures, and hypergraph partitions. In *Proceedings of the 6th ACM International Conference on Web Engineering*, pages 48–55, 2006.
- [20] C.-W. Ngo, W.-L. Zhao, and Y.-G. Jiang. Fast tracking of near-duplicate keyframes in broadcast domain with transitivity propagation. In *Proceedings of the 14th ACM International Conference on Multimedia*, pages 845–854, 2006.
- [21] K. Selvi and A. Tamilarasi. Association rule mining with dynamic adaptive support thresholds for associative classification. In *Proceedings of the IEEE International Conference on Computational Intelligence and Multimedia Applications*, volume 2, pages 76–80, 2007.
- [22] M.-L. Shyu, S.-C. Chen, and R. L. Kashyap. Generalized affinity-based association rule mining for multimedia database queries. *Knowledge and Information Systems (KAIS), An International Journal*, 3(3):319–337, 2001.
- [23] S. Tan, H.-K. Tan, and C.-W. Ngo. Topical summarization of web videos by visual-text time-dependent alignment. In *Proceedings of the ACM International Conference on Multimedia*, pages 1095–1098, 2010.
- [24] M. Wang, Q. Zou, and C. Liu. Multi-dimension association rule mining based on adaptive genetic algorithm. In *Proceedings of the IEEE International Conference on Uncertainty Reasoning and Knowledge Engineering*, volume 2, pages 150–153, 2011.
- [25] X. Wang, C. Zhai, X. Hu, and R. Sproat. Mining correlated bursty topic patterns from coordinated text streams. In *Proceedings of the 13th ACM International Conference on Knowledge Discovery and Data Mining*, pages 784–793, 2007.
- [26] WEKA. <http://www.cs.waikato.ac.nz/ml/weka/>.
- [27] I. H. Witten, E. Frank, and M. A. Hall. Data mining: Practical machine learning tools and techniques, 3rd edition. 2011.
- [28] X. Wu, Y.-J. Lu, Q. Peng, and C.-W. Ngo. Mining event structures from web videos. *IEEE Multimedia*, 18(1):38–51, 2011.
- [29] X. Wu, C.-W. Ngo, and A. G. Hauptmann. Multimodal news story clustering with pairwise visual near-duplicate constraint. *IEEE Transactions on Multimedia*, 10(2):188–199, 2008.
- [30] X. Wu, C.-W. Ngo, and Q. Li. Threading and autodomumenting news videos: a promising solution to rapidly browse news topics. *IEEE Signal Processing Magazine*, 23(2):59–68, 2006.
- [31] X. Wu, W.-L. Zhao, and C.-W. Ngo. Near-duplicate keyframe retrieval with visual keywords and semantic context. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, pages 162–169, 2007.
- [32] J. Yao, B. Cui, Y. Huang, and Y. Zhou. Bursty event detection from collaborative tags. *World Wide Web*, 15(2):171–195, 2012.
- [33] B. Yu, W.-Y. Ma, K. Nahrstedt, and H.-J. Zhang. Video summarization based on user log enhanced link analysis. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, pages 382–391, 2003.
- [34] Y. Zhai and M. Shah. Tracking news stories across different sources. In *Proceedings of the 13th ACM International Conference on Multimedia*, pages 2–10, 2005.
- [35] D.-Q. Zhang and S.-F. Chang. Detecting image near-duplicate by stochastic attributed relational graph matching with learning. In *Proceedings of the 12th ACM International Conference on Multimedia*, pages 877–884, 2004.
- [36] W.-L. Zhao, X. Wu, and C.-W. Ngo. On the annotation of web videos by efficient near-duplicate search. *IEEE Transactions on Multimedia*, 12(5):448–461, 2010.