2015 IEEE International Conference on Computational Intelligence & Communication Technology

A Novel Approach for Video Classification Based on Association Rules

Lu Bo
Dalian Key Lab of Digital Technology for National Culture, Dalian Nationalities University,
Liaoning Dalian, China
lubo@dlnu.edu.cn

Duan Xiaodong*
Dalian Key Lab of Digital Technology for National Culture, Dalian Nationalities University,
Liaoning Dalian, China
*duanxd@dlnu.edu.cn

Abstract—**Video classification is an important task for processing, analysis and retrieval of videos. The traditional method of Video classification generally use HMM theoretic. However, the HMM method has limitation to analysis video data. To solve this problem, we proposed a novel approach for video classification which uses the association rules. Firstly, we mined the actual dependence relationship between video states when the state model is constructed. Secondly, the reliability of dependence relationship is predicted with the restriction of association distance. Furthermore, the association rules are obtained by exploring state transition patterns. The experiment results demonstrate that the proposed method is efficient and suitable for various types of video data.**

Key words-association rules; state transition pattern; association distance; video classification

I. INTRODUCTION

With the rapid development of multimedia and Internet technology, the amount of huge video collections is increasing at a tremendous speed. In order to use these information of video efficiently, people need to retrieval and index automatically the source of large scale video data. Then, the task of video classification is the first step of content-based video retrieval; it can be regard as the most important component of processing of video data [1].

The method of video classification can be divided into two categories: (a) the first class is easy to realize by considering the space features of video data, but the results of classification often lead to the errors of semantic [2]. (b) the second class considers both the features of space and tempore, which can obtain the satisfactory results of classification [3]. Since the tempore features of video data ensures the completeness of semantic level of video. Some multimedia research communities generally use the theory of Hidden Markov Model (HMM), because HMM can obtain the time statistics information of stochastic process [4] [5]. However, the HMM model has its own drawback to analysis and process video data, which result in the unsatisfactory results of video classification.

To solve above motioned issues, we proposed a novel approach of video classification based on Association Rules (AR). We firstly construct the states model of video data and acquire the relationship among states by mining the states symbol sequence. Then, the reliability of dependence relationship is estimated with restriction of association distance and the State Transition Patterns (STP) is obtained. Furthermore, the association rules which are constructed by state transition patterns are used to perform video classification. The proposed method solves the limitation of HMM in the following three aspects: (a) the number of states of association rules is not fixed; (b) association rules is obtained from the procedure of mining the video data, which not need to pre-training the model; (c) the data type corresponding to the different association rules, which improve the accuracy of the classification results. The experiment result shows that proposed method is applicable to the various types of video data, the performance of our approach is more efficient than HMM method.

II. PRECESSING OF VIDEO DATA

A. State Model of Video Data

Shot and scene is two important concepts for video data. The shot of video can be regard as the physical boundary of video data, which is the smallest unit besides of video frame. The scene can be regard as the semantic boundary of video data, which use to express a complete video content. In this paper, we use the shot of video as states of video data, and construct the state model. Because the implementation of detection of shot of video is easier than scene, and the shot of video tends to facilitate to analysis. Therefore, a unified detection of shot boundaries algorithm is used to segment the shot of video [6].

B. States Symbol Sequence

When the state model is constructed, the key frame is abstracted from the shot of video and each key frame corresponding to a state. Then, we use split-and-merge method to cluster the states [7]. Meanwhile, the symbols use to mark results of clustering, which mean to mark each state. Then, the video data can be transforming to the states symbol sequence.

III. MINING OF ASSOCIATION RULES

A. State Transition Patterns And Assocition Rules

Given video data set $V=\{v_1,v_2,...,v_n\}$, $C=\{c_1,c_2,...c_n\}$ indicates the class of video data. For each v_i, $C(v_i)$ indicates the class of v_i. Table I shows that a video data set belong to two categories. In order to formulate the

978-1-4799-6023-1/15 $31.00 © 2015 IEEE
DOI 10.1109/CICT.2015.161

definition of state transition patterns, we first give some definitions of related concepts as following.

TABLE I. VIDEO DATA SET IS REPRESENTED BY STATE SYMBOL SEQUENCE

Video	state symbol sequence	class
V_1	ABBCCAEDBADEEDBAD	C_1
V_2	ABHBGABDBHGDABDHG	C_2
V_3	BCBCAEDBADAEDBAA	C_3
V_4	HBGABBDBHGABDHG	C_4

Definition 1. Association Distance (AD). The association distance between different states is defined as the distance between state symbol sequences. For example, as shown in Table I, the association distance between the first state A and the second state B of v_1 is $AD(AB)=1$. Corresponding to the fourth state C, the association distance is 3. The greater of associate distance implies the relationship among state is greater. Therefore, the association distances $AD_{\max}$ can be regards as the threshold to constraint the relationship between states.

Definition 2. Association Patterns (AP). For the state symbol sequence, $S=\{s_1,s_2,...,s_n\}$ indicates the state symbols of different class in the sequence, s_i is a arbitrary state symbol, n is the number of class in the state symbol sequence. The combination of any state symbol $\{s_i,...,s_k\}$ is an association patterns.

Definition 3. Reliability of Association Patterns (RAP). The reliability of association patterns is a ratio between the numbers of times appears in the state symbol sequence and the total number of states. Such as, in the Table I, for the v_1, association distance $AD_{\max}=2$ appear three times, so the $RAP\{AD\}=3/17\approx 18\%$. The greater of RAP indicates that the ability of prediction of classification is more precise. Therefore, $RAP_{\min}$ can be regards as the threshold value for analysis the availability of association patterns.

Definition 4. State Transition Pattern (STP). Given the minimum of reliability of association patterns $RAP_{\min}$, all the value of *RAP* greater than $RAP_{\min}$ of association pattern which is called state transition pattern. Table II shows that a part of STP ($AD_{\max}$=2, $RAP_{\min}$=10%) corresponding to the Table I.

TABLE II AN EXAMPLE OF STATE TRANSITION PATTERN

Video	State transition pattern
V_1	{AD} {BD} {ED} {AED} {EBD}
V_2	{AB} {AD} {BH} {BG} {GA} {BHG}
V_3	{AD} {AE} {ED} {AEB} {AED}
V_4	{AB} {BD} {BH} {GA} {GB} {BDH}

Definition 5. Association Rules (AR). The association rules include the class of video data and state transition patterns, which is formulated as $STP_k \to C_i$.

Definition 6. Association Rules Support (ARS). The support of association rules $STP_k \to C_i$ is defined as a ratio between the number of video data which include the STP_k and the total number of STP of entire video data sets. The association rules support (ARS) is defined as:

$$ARS(STP_k \to C_i)=\frac{\sum_{C(v_j)}(v_j,STP_k)}{\sum_{j=1}^{|V|}N(v_j)} \qquad (1)$$

Where (v_j,STP_k) indicates the video data v_j include the state transition patterns STP_k, $N(v_j)$ denote the number of *STP* of v_j, $|V|$ indicates the number of video data sets.

Definition 7. Association Rules Confidence (ARC). The confidence of association rules $STP_k \to C_i$ is defined as a ratio between the numbers of video which includes STP_k with C_i and the total number of video data sets which includes STP_k. We formulate ARC as following:

$$ARC(STP_k \to C_i)=\frac{\sum_{C(v_j)=C_i}(v_j,STP_k)}{\sum_{C(v_j)}(v_j,STP_k)} \qquad (2)$$

B. 2-Hash Data Structure

In this section, the ARS and ARC of association rules is obtained by 2-Hash data structure. Because 2-Hash data structure just need travel over video data sets by one time, and save the *I/O* cost to a large extent.

2-Hash is a two levels structure of hash table. The first level includes a hash table, which aim to hash the first symbol in the *STP*. The second level includes some hash tables, which aim to hash the last symbol in the *STP*. Moreover, the record of the second level hash table is a list structure, each record consists of two part information, the first part is called symbol attribute, which denotes the remaining symbol except of the first symbol and the last symbol in the state transition pattern. The second part is called statistics attribute, which denotes the number of class of video data with including *STP*. It is formulated by $[N(C_1),N(C_2),...,N(C_n)]$, where $N(C_n)$ indicates the number of video data including STP belong to the class C_n, the initial value of statistics attribute is set to 0. The record of list is ranked by symbol attribute for facilitating the query.

The procedure of 2-hash structure need travel over the video data sets by one time. For a *STP*, the first symbol of *STP* as a keyword which is hashed to the first level hash table. Then, the last symbol in the *STP* is hashed into the second level hash table by corresponding to the keyword.

Finally, the symbol attributes and statistics attributes is inserted into the hash table, and the total statistics information of 2-hash structure is modified. For example, given $\{AD\}$ of v_1, the symbol A is hashed into the first level hash table, and the symbol D is hashed into the second level hash table. Furthermore, the corresponding record is inserted into the record level.

The support and confidence of association rules is obtained by 2-hash structure. Such as the state transition pattern $\{AD\}$, we need to find the statistics information of $\{AD\}$ in the 2-hash structure for acquiring the *ARS* and *ARC* of $\{AD\}$. We assume [12,14] as the total statistics information of 2-hash structure. According to the Definition 6 and 7, the *ARS* and *ARC* of the association rules $AD \to C_1$ and $AD \to C_2$ can be estimated as following:

$ARS(AD \to C_1) = 7.69\%$, $ARC(AD \to C_1) = 66.7\%$, $ARS(AD \to C_2) = 3.84\%$, $ARC(AD \to C_2) = 33.3\%$.

IV. VIDEO CLASSIFICATION BASED ON ASSOCIATION RULES

A. Training

The procedure of training is to find association rules in our proposed method. Meanwhile, the ARS and ARC of association rules are used to classify the video data sets. The entire procedure of training consists of two sections: (a) the video data set is firstly pre-processed, and transformed into the sate symbol sequence. (b) the state symbol sequence is mine for obtaining the state transition patterns, and construct the 2-hased structure.

B. Classification

According to the above motioned, the video testing data sets need to be preprocessed along with the training data set for obtaining the *STP* of test data sets. Then, the relative information of association rules in the training phase is used to classify the video test data sets.

For example, given a video data sets V_t, we need firstly to query every *STP* in the 2-Hash structure, and calculated the *ARS* and *ARC* of association rules which consisted by *STP*. Moreover, we explore a simple and efficient method to prune the association rules, the aim of pruning is to delete the little reliability of association rules for improve the accuracy of prediction and classification.

Meanwhile, the proposed method is not considering the association rules by training data sets, but the reliability of *STP* of testing data. Therefore, we formulate the results of classification by average reliability degree as following:

$$\bar{C}(V_t, C_k) = \frac{\sum_{STP_i \in P_t} RAP(STP_i) \times ARS(STP_i \to C_k)}{\sum_{STP_i \in P_t} RAP(STP_i)} \quad (3)$$

Where $\bar{C}(V_t, C_k)$ indicates the average reliability degree of video data V_t belong to C_k. P_t is the *STP* sets of V_t, the $RAP(STP_i)$ denotes the reliability degree of STP_i.

V. EXPERIMENTAL EVALUATION

In this section, we evaluate the performance of our proposed approach and introduce the setup of the experiments, including the data preparation, experimental environment and parameter setting.

The video data sets include 80 video clips with MPEG format. All the video clips derive from 4 different data type: *News* (*N*), *Basketball* (*B*), *Tennis ball* (*T*) and *Football* (*F*). The entire video data sets are divided into training set and testing set. The performance of classification is evaluated by precision and recall. All the experiments are conducted on Intel Core 4 2.8 GHz CPU with 4GB memory and a 500GB hard disk.

We select 40 video clips as training set, and obtain 244 state transition patters. The *ARS* and *ARC* of association rules corresponding to a part of STP is demonstrated in Table III.

TABLE III THE *ARS* AND *ARC* OF ASSOCIATION RULES OF TESTING SETS

AR	ARS(%)	ARC(%)
$\{AD\} \to C_1$	5.88	83.33
$\{GBD\} \to C_2$	4.71	62.5
$\{GBDH\} \to C_2$	5.88	95.6

As shown in Table III, our proposed method is better than HMM, the advantage of our method mainly reflect in three aspects: (a) the number of state of HMM is fixed, but the number of *STP* is not fixed. (b) a kind of data in HMM model is mapped into one *STP*, but our method can mapped into multi-association rules. (c) the HMM model need to a procedure of pre-learning phase to assure the number of state and *STP*, but our method directly obtain the *STP* and association rules.

The results of precision and recall is shown in Figure 1 and Figure 2, when the given $ARS_{\min} = 2\%$, $ARC_{\min} = 50\%$ and $ARS_{\min} = 4\%$, $ARC_{\min} = 50\%$. We observe that the results of precision and recall are declining with the value of $ARS_{\min}$ increasing, but the accuracy of classification is still having a high performance. Because the reliability of association rules is enhanced binding by $ARS_{\min}$. This situation leads to the association rules become no longer reliable. Therefore, the average reliability degree is declining when the number of valid association rules is reduced. The performance of classification is the best with $ARS_{\min} = 2\%$ by the repeated experiment.

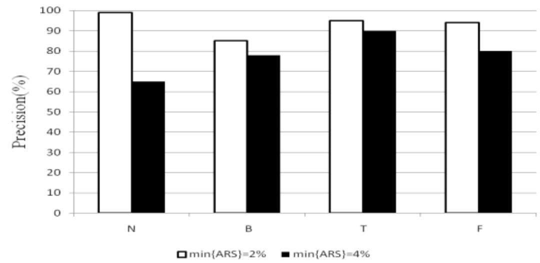

FIGURE 1. COMPARISON OF PRECISION WITH ASSOCIATION RULES

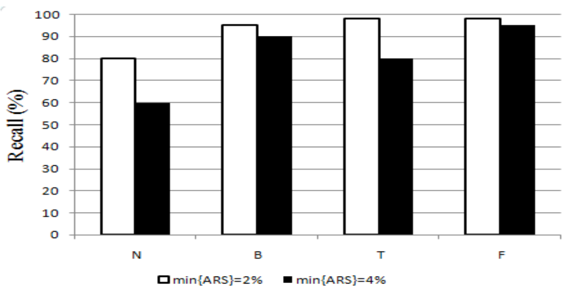

FIGURE 2. COMPARISON OF RECALL WITH ASSOCIATION RULES

The performance of our proposed method and HMM method is compared as shown in Figure 3 and Figure 4. We train the HMM model by exploring the same training set and predict the result of classify by using the same testing set. In our proposed method, we set the $ARS_{\min} = 2\%$ and $ARC_{\min} = 50\%$, the number of states of HMM method is 3. We observed that the performance of our approach is better than HMM method, especially, the precision of classification is improved greatly for the sports video. The main reason is that the sports video is relatively fixed, therefore, the association rules is more reliable, which lead to the high performance of video classification.

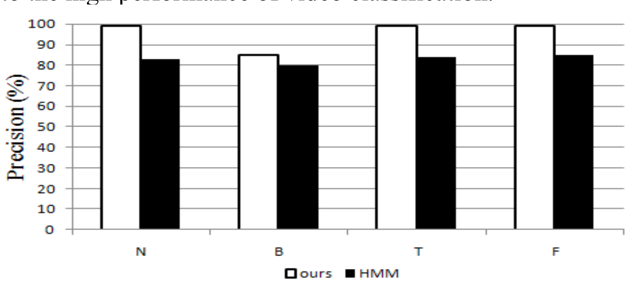

FIGURE 3. COMPARISON OF PRECISION WITH OUR METHOD AND HMM

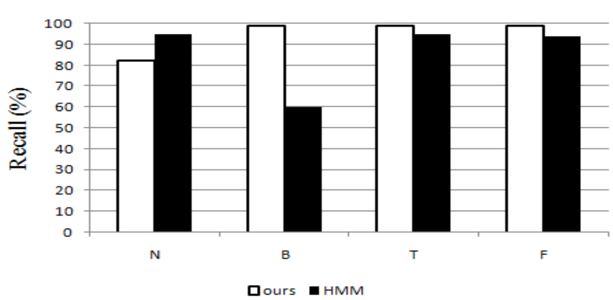

FIGURE 4. COMPARISION OF RECALL WITH OUR METHOD AND HMM

VI. CONCLUSIONS

In this paper, we proposed a novel and efficient method based on association rules for improving the problem of video classification. This method simultaneously considers both space feature and temporal feature of video data. The prediction of classification is conduct by using association rules. The experiment results demonstrate that the proposed method is efficient and suitable for various types of video data.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant No.61370146; National Natural Science Foundation of China under Grant No.2013405003; Fundamental Research Funds for the Central Universities.

REFERENCES

[1] X. Q. Zhu, X. D. Wu, A. K. Elmagarmid. Video data mining: Semantic indexing and event detection from the association perspective. IEEE Trans on Knowledge and Data Engineering, 2005, 17(5): 665-677.

[2] W. Zhou, A Vellaikal. Rule-based video classification system for basketball video index. Proceedings of ACM Workshops on Multimedia, ACM Press, 2000, pp. 213-216.

[3] G Wei, L Agnihortri, N. Dimitrova. TV program classification based on face and text processing. Proceedings of IEEE Multimedia and Expo, 2000, 7(3): 1345-1348.

[4] C. Lu, M. S. Drew, An automatic video classification system based on a combination of HMM and video summarization. Journal of Smart Engineering System Design, 2003, 5(1): 33-45.

[5] E Kijak, G. Gravier, P. Gros. HMM based structuring of tennis videos using visual and audio cues. Proceeding of Multimedia and Expo, 2003, pp. 105-111.

[6] X. Q. Zhu, J. P Fan, A. K. Elmagarmid. Hierarchical video content description and summarization using unified semantic and visual similarity. Journal of Multimedia Systems, 2003, 9(1): 31-53.

[7] H. Yi, D. Rajan, L. Chia. A unified approach to detection of shot boundaries and subshots in compressed video. Proceedings of Image Proceesing. 2003, pp. 56-64.