Video Key Frame Extracting Scheme Based on Layered Optimizing SVM Algorithm

Haiming Cui  
Langfang Teachers’ College, Langfang 065000, Hebei, China

Dazhi Wang  
Langfang Vocational and Technical College, Langfang 065000, Hebei, China

Abstract  
In order to overcome the low accuracy of standard SVM (support vector machine) algorithm in extracting video key frame, a model of extracting video key frame is proposed based on weak hierarchical optimizing SVM algorithm. First, the final classifier is generated according to calculating deviation of each SVM classifier to adjust the weights of the SVM classifier by BOOST algorithm, and video frames are for weak classification. And then the output of the SVM classifier is for decision fusion by D - S evidence theory, and the parameters of Sigmoid function are used to fitting a posteriori probability of SVM output directly, and the video frames are for strong classification. Simulations show that compared with the standard SVM algorithm, the proposed model can extract the corresponding key frames with higher accuracy.

Key words: SVM Classification Model, Video Analysis, Key Frame Extraction, Hierarchical Optimization, D - S Rules, The BOOST Algorithm.

1. INTRODUCTION

In recent years, all kinds of video information are growing rapidly, such growth trends forcing people to develop video retrieval and classification techniques. Because people use high-level semantic concepts to query and browse the video database, so it is necessary to develop the semantic analysis technology of video content with automatic and efficient visual information retrieval and browsing (Wu, Ma and He, 2013). Video usually longer, on the other hand, when people want to skip a boring part directly and to watch the highlights, will require the system to help users choose the wonderful clips and summary. Researching video classification retrieval can help users to quickly retrieve efficiently to the needs of content (Xiao and Chun, 2011).

With the improvement of information technology, video classification research has made considerable development in recent decades, the domestic and foreign proposed many methods and models in the field of video classification. Noboru Babaguchi et al. using PCA principal component to reduce the dimensions of the video and audio visual features to describe the video content, and then use the motion characteristics of the time sequence to distinguish action events classification in the sports video (Lin, Liu and Wang, 2012). According to the characteristics such as the field area and the field distribution, Ekin and others implement the classification of video camera, event detection and generate video abstraction (Smeulders, Worring and Santini, 2010). Through some movement pattern detection in video frames, such as the lens shifts like running, jumping, serve, and scaling and so on, Ma and others classify the simple movement in video (Zhang, Cao and Zhou, 2013). Xu Rong presents a video retrieval algorithm based on the characteristics of audio, which can mark specific events such as a score in the video, and slow motion playback (Xie, Xu, Chang and Divakaran, 2014). David Liu achieves the classification of video by separating out the video information such as the target, and the effect is good when the video background is unitary and shade the area between the objects a few cases, but results is not very well when the video overlap between background complex and target is large (Hu and Zhao, 2013). According to the structured video data, Lin Bin detect and locate the slow motion in video by logo template matching first, then calculate the pixel ratio of each sub-block in field in normal video camera frame based on the idea of blocking, as the feature input SVM classifier while implements segmentation of the game shot, but the object is limited to the soccer video with insufficient popularization (Manjunath and Rainerohm, 2014). Ba Tu Truong extracts video editing, color, motion feature, and implements the classification of the video by constructing a decision tree, but because of the restriction of the decision tree, the results often convergence in local optimal solution, have an error (Hu, Zhao and Chen, 2014). Xavier Gibert et al. extract feature vectors which are formed by video motion and the main color information, establish a classifier based on HMM model and video is divided into various projects, the experimental accuracy is higher, but the HMM is used as classifier, requires a large number of training samples and the observed sequence, which causes the increase of calculation (Snoek and Worring, 2009). Geeth proposes a video classification method based on HMM,
implement the classification of the video through the selection of relatively simple features. And its classification model often rely on a large number of training data, has led to an increased workload greatly in practical work, in addition, the observed sequence of the HMM model also need to be long enough, and calculation also increased (Zeng and Xu, 2013). Zhang Longfei and others detect the characters in the video and the sports area by using edge detection and mathematical morphology method, through the information such as the environment scene and color of character and area ratio and the SVM classifier to classify video (Lin, Lin and Weng, 2014).

According to the defects of the standard SVM algorithm in extraction of key frame, a model of extracting video key frame is proposed based on strength and weak layered optimized SVM algorithm, and simulations show that the improvement strategy is valid.

2. DEFECT ANALYSIS OF KEY FRAME EXTRACTION BASED ON SVM ALGORITHM

Support vector machine (SVM) is based on the principle of structural risk minimization, which can better solve learning problems in the traditional learning algorithm, and the SVM algorithm has higher classification precision and efficiency.

A training set is given as follows.

\[ T = \{(x_i, y_i),..., (x_l, y_l)\} \in (R^n \times y)^l \]  

In the equation (1): \( x_i \in R^n, y_i \in y = \{-1, 1\}, i = 1,...,l \). The root of the classification is to find a real function \( g(x) \) on the space \( R^n \), through the decision function as the equation (2) to concluded that the output corresponds to any input \( x \).

\[ f(x) = \text{sgn}(g(x)) \]  

Consider the classification problem of two-dimensional space \( R^2 \). If a straight line can divide the space \( R^2 \) into two parts, obviously there are many lines can divide the two types of points and we need to find the better line.

By calculating the distance between two straight lines can get the distance of two straight lines as \( \frac{2}{\|w\|} \), to maximize the ideas of “interval” classification problem is transformed into the optimal problem of solving variables \( w \) and \( b \).

\[ \min_{w,b} \frac{1}{2}\|w\|^2 \]  

The dual problem: solve the biggest interval problem can not directly to solve the optimal problem (3), to solve the optimal problem of the dual problem need to introduce Lagrange function.

\[ L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum a_i(y_i((w \cdot x_i) + b) - 1) \]  

In the equation (4), \( a = (a_i,...,a_l)^T \) are multiplier vectors of Lagrange. Optimal problem is shown as equation (5).

\[ \max_a -\frac{1}{2}\sum_{i=1}^l \sum_{j=1}^l y_i y_j (x_i \cdot x_j) a_i a_j + \sum_{i=1}^l a_i \]  

Lagrange function found that not all classification problem can be obtained by linear classification, and there may not be such a hyperplane, therefore the starting point is usually doesn’t work. Obviously we have to “soften” the request for cutting hyperplane, so we can continue to insist on a hyperplane partition. “Soften” is the pools by introducing loose variables, while existing training points that not satisfy the constraint conditions \( y_i((w \cdot x_i) + b) \geq 1 \).

\[ \xi_i \geq 0, i = 1,2,...,l \]  

The constraint conditions of “soften” are shown as follows.

\[ y_i((w \cdot x_i) + b) \geq 1 - \xi_i \]
Of course, when $\xi_i$ is sufficiently large, training point $(x_i, y_i)$ can always satisfy the above constraints. In order to constraint value of $\xi_i$, the solution is making $\sum_i \xi_i$ join in the optimization function. $\sum_i \xi_i$ plays a role of punishment. The original optimization problem is instead of the original problem.

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i$$

(8)

In the equation (8) $\xi = (\xi_1, \ldots, \xi_n)^T$, $C > 0$ is the parameter for punishment. The Lagrange function is introduced in the equation (9).

$$L(w,b,\xi,\alpha,\beta) = \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i (y_i ((w \cdot x) + b) - 1 + \xi_i) - \sum_{j=1}^{m} \beta_j \xi_j$$

(9)

$\alpha = (\alpha_1, \ldots, \alpha_n)^T$ and $\beta = (\beta_1, \ldots, \beta_m)^T$ are Lagrange multipliers vectors.

Optimization problem is the dual problem of the original problem.

$$\max_{\alpha,\beta} - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) + \sum_{j=1}^{m} \beta_j$$

(10)

Some classification problems are unfavorable to adopt linear partition, and should adopt with nonlinear division. The key is introduced an appropriate transformation $\Phi$. Assume the original training set as equation (11).

$$T = \{(x_i, y_i), i = 1,\ldots,l\} \in (R^n \times \gamma)^l$$

(11)

In the equation (11), $x_i \in R^n, y_i \in \gamma = \{-1, 1\}$, and introduce the transformation $x = \Phi(x)$ from space $R^n$ to the Hilbert space $H$:

$$\Phi : R^n \rightarrow H$$

$$x \rightarrow x = \Phi(x)$$

(12)

Determine the optimal classification in SVM is only a small number of the samples of support vectors, so the SVM has the strong ability to adapt small sample data set. In addition, as a convex optimization problem, the local optimal solution of SVM is surely a global optimal solution. Considering the many advantages of SVM, we adapt the SVM as classifier of video key frames.

But the existing algorithm is difficult to achieve a satisfactory in time and space. The main reason for the low efficiency of training algorithm is traditional SVM to adopt the standard quadratic optimization technology to solve the dual problem. First, since each iteration of SVM needs to save the whole kernel function matrix in computer memory, when the number of training sample is large, which causes the consumption of SVM training memory increased dramatically. Second, the SVM training algorithm calculation process is an iterative process, this led to the long training time and consume memory of SVM algorithm.

3. SVM ALGORITHM BASED ON HIERARCHICAL OPTIMIZATION

3.1. Weak Classification Based on BOOST Algorithm

Combined with the BOOST algorithm, we introduce the concept of weights in the SVM. Unlike the BOOST algorithm in this paper, not by adjusting the weights of weak classifier to produce the final classifier, but adjust the weight of SVM classifier according to the deviation of each SVM classifier, and generate the final classifier.

The equation of calculation deviation is as follows.

$$\varepsilon_i \leftarrow \sum_{i=1}^{n} |D_i(x_i) - y_i|$$

(13)

$D_i$ is classifier, $x_i$ is video vector, $y_i$ for the classification information of video with the value of 0, 1.

According to the feedback of samples obtained from training the classifier save as $D_i$, and then calculate the deviation $y_i$ of each $D_i$, given $D_i$ different weights according to the equation (14) and finally form the classifier.
According to the characteristics of key frame extraction at the same time, the user’s positive feedback has similarities, while negative feedback has no clear similarities, so we adjust sample set of weak classifier besides classifier \( D_1 \). All positive samples are joined in sample set, and the negative samples adopt the results by negative feedback, and get a positive feedback in common as far as possible. So the samples set \( S_i \) is as follows.

\[
S_i = L_i \cup S_{i-1} \cup \{ X : X \in U, y = 1 \}
\]

The specific algorithm process is as follows:

BEGIN
(1) Initialize the sample set \( U \), made \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\) whether belong to the label set \( y_i = 0, 1 \);
(2) Initialize the deviation of standard classifier \( \epsilon_i \leftarrow \sum_{i=1}^{n}[D_i(x_i) - y_i] \);
(3) Train classifier:
   1) Train \( D_i \) according to \( S_i \);
   2) Update classifier weight according to equation (14);
   3) The final classifier is shown as follows.
   \[
   C = \sum_{i=1}^{k} w_i \cdot D_i(x_n)
   \]
(4) Sort the results;
(5) IF users are not satisfied with the results THEN
   Calculate \( S_j = L_k \cup S_{k-1} \cup \{ X : X \in U, y = 1 \} \);
   GOTO training classifier;
ELSE
   The end;
ENDIF
END

3.2. Strong Classification Based on D - S Rules

Dempster - Shafer (D - S) evidence rule is a law reflects the combination between the evidence. It can merge two or more trust function (Bel), through calculate the trust of different sources of orthogonal and find a new trust function. Synthesis rule of Dempster merging various trust function is through the basic probability distribution function, the trust function is defined in the same recognition framework, but the requirements for evidences which is independent, in the theory of evidence combination rules, the independence of the evidence is the crucial factor.

Given several basic probability assignment BPA based on the same recognition framework with different evidence, if the evidence not completely conflict, then we can merge out a total BPA by Dempster synthesis rules that is the result of combining some evidence. This function is the orthogonal original BPA, marked as \( \oplus \).

Dempster fusion rules: set \( m_1, m_2, \ldots, K \) and \( m_n \) as basic distributive function of different evidence on the same recognition framework \( \Theta \), their orthogonal sum \( m = m_1 \oplus m_2 \oplus K \oplus m_n \) is determined by the equation (17).

\[
m(A) = (m_1 \oplus m_2 \oplus A \oplus m_n)(A)
\]

In the equation (17),

\[
K = \sum_{A \cap A_1 \cap A_2 \ldots \cap A_n = \emptyset} m_1(A_1) \cdot m_2(A_2) \cdots m_n(A_n)
\]

When \( A = \emptyset \), \( m(A) = 0 \).

When \( K \neq 1 \), \( m_1 \oplus m_2 \oplus A \oplus m_n \) is still the basic probability distribution function after combined; when \( K = 1 \), there is meaningless and contradictory between every \( m_n \).

In order to use the D - S evidence theory to decision fusion for the output of the SVM classifier, it needs to obtain the basic probability assignment from the output of single classifier, Sigmoid function with parameter is

\[
w_k = \frac{e_k}{\sum_i e_k}
\]
used to directly fitting the SVM output of posterior probability, we estimate the parameters of the Sigmoid A and B, instead of directly estimate the posterior probability. In the two classification problem, the probability output form of the SVM is as equation (19).

\[
p(y = 1| f(x)) = \frac{1}{1 + \exp(Af(x) + B)} \\
p(y = -1| f(x)) = 1 - \frac{1}{1 + \exp(Af(x) + B)}
\]  

(19)

In the equation (19), the parameters A and B control the shape of the Sigmoid function, \( f(x) \) is the distance between the unknown test \( x \) to the optimal hyperplane. As long as \( A < 0 \), monotonicity is sure in the equation (19). For solving the parameters \( A \) and \( B \), we can get the optimal parameters through minimize training the negative logarithm likelihood value of data, which are defined as follows.

\[
\min\left( \sum_i (\log(p_i) + (1 - t_i)\log(1 - p_i)) \right)
\]  

(20)

In the equation (20),

\[
p_i = \frac{1}{1 + \exp(Af_i(x) + B)}
\]  

(21)

\[
t_i = \begin{cases} 
    \frac{N_i + 1}{N_i + 2}, & y_i = +1 \\
    \frac{1}{N_i + 2}, & y_i = -1 
\end{cases}
\]  

(22)

\( l \) is the number of training samples, the training set \((f_i(x), y_i)\) can be defined for the new training set \((f(x), t_i)\) of minimum optimization problem, and then use the generic optimization method can get the parameters \( A \) and \( B \).

In D - S evidence, the basic probability assignment (BPA) represents the support of a evidence for a particular example, is the basic input information of evidence synthesis, in target fusion based on evidence theory, BPA acquisition is vital, and basic probability assignment method is whether appropriate or not, will directly affect the effectiveness and accuracy of decision fusion results.

With the posteriori probability output \( p_i \) of SVM, we can complete the structure of the BPA. For any one of two classes of the SVM classifier, after finishing the study of sample set, a best SVM model is obtained by optimized, get the optimal parameters \( A \) and \( B \) according to the equation (17), and thus constructing a posteriori probability model \( M_r \), and then testing retention validation set, computing the performance of each probability SVM, get the accuracy rate \( r_i \) of each probability SVM classification. Among them \( r_i \in [0,1](i = 1,2,\ldots,k) \), \( k \) is the number of classifier. If retain the capacity \( c \) of validation data set, the \( i \) SVM recognize \( q_i \) samples, while classification accuracy of \( i \) probability SVM is as follows.

\[
r_i = \frac{q_i}{c}
\]  

(23)

Through the above data can get the basic probability assignment BPA of each probability SVM about D - S evidence theory. If the classification accuracy rate of a probability SVM is \( r \), this suggests that the probability of a unknown attribute of sample can be identified by the probability SVM is \( r \), namely the probability of SVM classifier can not determine the input unknown sample belongs to one class is \( 1 - r \), so here’s the uncertain probability \( 1 - r \) of SVM classifier is appropriate to the uncertain information in the D - S evidence theory, \( 1 - r \) can be regarded as the uncertain information of probability SVM classifier in fusion decision. So, in the basic probability distribution, the classification accuracy \( 1 - r \) of uncertainty is assigned to identify framework that is reasonable, \( m(\Theta) = 1 - r \), and get three classification results: the positive probability, the negative probability and the uncertain probability. Assuming that the probability output and classification accuracy of the \( i \) probability SVM classifier is as \( p_i \) and \( r_i \) respectively, the three classification results respectively as three incompatible elements of recognition framework, using the concept of the above analysis and evidence theory, we define a basic probability function of binary SVM.
In the equations, \( m_d(x) \) is the degree of uncertainty, we can get positive class probability and the negative class probability as well as the uncertainty probability of samples \( x \) according to the above equations.

### 4. PERFORMANCE SIMULATION OF ALGORITHM

In order to verify the effectiveness of the proposed improved algorithm in this paper, move on to the simulations. Taking two surveillance videos, adopt the standard SVM algorithm and the proposed improved SVM algorithm on the “theft” and “loot” key frame extraction, the results are shown in the figures.

**Figure 1.** Result of stealing video key frame extraction based on standard algorithm

**Figure 2.** Result of stealing video key frame extraction based on improved algorithm
Figure 3. Result of loot video key frame extraction based on standard algorithm

Figure 4. Result of loot video key frame extraction based on improved algorithm

Seen from the simulations, the proposed model of extracting video key frame is very good to extract the corresponding key frames based on strong and weak layered optimized SVM algorithm, and the accuracy of the standard SVM algorithm to extract key frames is not enough.

5. CONCLUSIONS

Now, watching video programs has become a kind of modern life and entertainment. As the rapid expansion of video shows, how to help users find their favourite sections is become urgency. According to the defects of the standard SVM algorithm in key frame extraction, a model of extracting video key frame is proposed based on strong and weak layered optimized SVM algorithm, simulations show that compared with the standard algorithm, the proposed algorithm has better accuracy in extracting video key frame.

REFERENCES


