ISSN:0975-9646

P.Thirumurugan, et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3473-3475

Event Detection in Videos Using Data Mining Techniques

P.Thirumurugan, S.Hasan Hussain

Syed Ammal Engineering college,Ramanathapuram

Abstract— Event detection is one of the essential tasks by which the performance of video content analysis and access becomes more efficient and effective. In this paper, the rare event detection issue in video event detection is addressed through the proposed data mining framework which can be generalized to be domain independent. The fully automatic process via the combination of distance based and rule-based data mining techniques can greatly reduce the number of negative (non-event) instances and the feature dimension to facilitate the final event detection, without pruning away any positive (event) testing instance along the process. The effectiveness and efficiency of the proposed framework are demonstrated over the goal event detection application based on a large collection of soccer videos with different styles.

***Keywords*— Event detection, distance based, rule based.**

I. Introduction

DATAMINING techniques have been increasingly developed to provide solutions for semantic event detection in diverse types of videos [6]. Video events are normally defined as the interesting events which capture user attentions. For example, a soccer goal event is defined as the ball passing over the goal line without touching the goal posts and the crossbar. Most current research for video event detection heavily depends on certain artifacts such as domain knowledge and priori model, and thus making them hard to be extendible to other domains or even other data sets. Though some work has been conducted to deal with the general video event extraction, they can only achieve rough detection capability [4] due to the well-known semantic gap and rare event detection issues [7].

In particular, the rare event detection issue, also known as imbalance data set problem, occurs in most video event detection applications. This issue is referred to as a very small percentage of positive instances versus negative instances, where the negative instances dominate the detection model training process, resulting in the degradation of the detection performance. In order to develop a generalized event detection framework applicable to different application domains, a necessary step is to relax the need of domain knowledge such as those domain-specific rules with pre-defined fixed thresholds and additional domain-based high-level features, which are often used to increase the percentage of the positive instances in the data set.

II. Related work

A number of recent research works have been focused on sports and news videos for event detection. Sports video analysis, especially sports events detection, has received a great deal of attention [1] owing to its great commercial potentials. For video content processing, many earlier studies adopted unimodal approaches that studied the respective role of visual, audio, and texture mode in the corresponding domain. Recently, the multimodal approach attracts growing attention as it captures the video content in a more comprehensive manner. In [10], a multimodal framework using combined audio, visual, and textual features was proposed.

In the decision-making stage, data mining has been increasingly adopted. For instance, [12] proposed a hybrid classification method called CBROA which integrates the decision tree and association rule mining methods in an adaptive manner. However, its performance is restricted by a segmentation process and a pre-defined confidence threshold. As far as video semantic analysis is concerned, support vector machines (SVM) are a well-known algorithm adopted for event detection [3] in sports videos and concept extraction [1] in TRECVID videos. Although SVM presents promising generalization performance, its training process does not scale well as the size of the training data increases [2]. C4.5 [5] is a matured representative data mining method, which was also applied in sports video analysis [9].

In this paper, the proposed framework attempts to achieve a fully automatic video event detection procedure via the combination of distance based and rule-based data mining techniques, which are two basic and widely used measurements in data mining. The proposed data mining framework is validated using soccer goal event detection as the testbed, and the experimental results on 2 soccer videos collected from different broadcasters demonstrate the powerfulness and potential of integrating distance-based and rule-based data mining techniques.

III. Proposed Framework

The proposed architecture is shown in Figure 1.It consists of Video Parsing and Feature Extraction, Distance-based Data Mining, and Rule-based Data Mining phases

A. Video Parsing and Feature Extraction

In soccer goal event detection testbed, raw soccer video is parsed via a video shot detection subcomponent and multilevel features are extracted based on multimodal theory [7]. Here, the feature set *F* contains totally 17 features including 10 audio features, 5 visual features, and 2 temporal features.

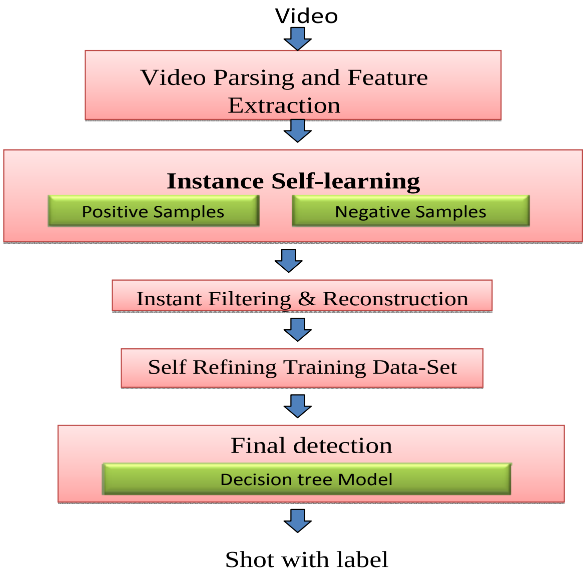

Figure. 1. Architecture of the proposed framework.

B. *Distance Based Data Mining*

The proposed distance-based data mining scheme is motivated by the powerfulness and robustness of our novel distance based anomaly detection algorithm called Representative Subspace Projection Modeling (RSPM) [11] under different application domains and diverse types of data sets.

a. Typical Instance Self-learning with Feedback

The proposed **typical positive instances** are defined as selected $T\ 1$ ($T\ 1 < N1$) positive instances in $\mathbf{X}^e = \{\mathbf{x}ij\}$. The normalized matrix $\mathbf{Z}^e=\{\mathbf{Z}ij\}$ of $\mathbf{X}^e$ can be obtained via Equation (1), where $\bar{\mu}_i$ and s_{ii} are the sample mean and variance of the *ith* row in $\mathbf{X}^e$.

$$Z_{ij} = \frac{X_{ij} - \bar{\mu}_i}{\sqrt{s_{ii}}} \qquad (1)$$

The principal components corresponding core row vectors $R^e_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \ ,\mathbf{y}_{iT\,1})$ of $\mathbf{Y}^e$ satisfying Equation (2) are selected as the representative components to model the similarity of the typical positive instances.

$$STD\ (R^e_m) < a\ , \qquad (2)$$

Where STD ($\mathbf{R}^e_m$) is the standard deviation of the score row vectors satisfying Equation (2) and a is the arithmetic mean of the standard deviation values from all $\mathbf{R}^e_i$ .Next, a class-deviation equation (Equation (3)) is designed to differentiate the normal and anomaly instances in the view of typical positive instances, where M is the selected index vector from Equation (2).

$$c^e_j = \sum_{m \in M} \frac{(y_{mj})^2}{\lambda^e_m} \qquad (3)$$

The maximum value $\mathbf{c}^e$ *max* of all $\mathbf{c}^e_j$ for $\mathbf{X}^e$ is selected as a threshold to justify if an incoming instance is statistically normal to the typical positive instances. Accordingly, in order to identify the best group $\mathbf{X}^e$, we randomly select $T\ 1$ instances from $\mathbf{X}^1$ for times. Each time, the percentage of the recognized instances in $\mathbf{X}^1$ and the percentage of the rejected instances in $\mathbf{X}^2$ are recorded, and $\mathbf{X}^e$ is defined as the one that can recognize 100% instances in $\mathbf{X}^1$ and at the same time reject the maximal percentage of instances in $\mathbf{X}^2$, when K and $T\ 1$ are big enough.Similarly, the proposed **typical negative instances** $\mathbf{X}^n$ is defined as the selected $T\ 2$ ($T\ 2 < N2$) negative instances which can reject 100% instances in $\mathbf{X}^1$ and at the same time recognize the maximal percentage of instances in $\mathbf{X}^2$.

b. Instance Filtering and Feature Reconstruction & Selection

Before the execution of this step, the original data set $\mathbf{X}$ is randomly split into two disjoint subsets, namely a training data set $\mathbf{X}^A$ (with known class labels) and a testing data set $\mathbf{X}^B$ (with unknown class labels).Assume that the labeled positive instances in $\mathbf{X}^A$ are $\mathbf{X}^{Ae}$ and the labeled negative instances are $\mathbf{X}^{An}$. From this step, data mining event detection is gradually achieved via distance-based filtering followed by rule-based classification. In this step, $\mathbf{X}^e$ and $\mathbf{X}^n$ are trained sequentially to conduct a rough classification to increase the percentage of positive instances, which addresses the *rare event detection* issue. $\mathbf{X}^e$ is first trained to reject negative instances in $\mathbf{X}^{An}$ and $\mathbf{X}^B$. The recognized normal data instances are passed to the second classifier trained with the selected $\mathbf{X}^n$. The original feature set F are reconstructed and filtered to be F' with the dimension p' ($p' < p$). The remaining and reconstructed data matrix can be defined as $\mathbf{Y}^{Ae}=\{\mathbf{Y}_{ij}\}$ for positive instances in the training data set, $\mathbf{Y}^{An}=\{\mathbf{Y}_{ij}\}$ for negative instances in the training data set, and $\mathbf{Y}^B=\{\mathbf{Y}^{ij}\}$ for all testing data where (i = 1, 2, . . . , p').

c. Self-refining Training Data Set

The self-refining of the training data set is achieved by linear analysis on the score row vectors $\mathbf{R}_1^{Ae}$ and $\mathbf{R}_1^{An}$ which correspond to the first selected principal component for $\mathbf{Y}^{Ae}$ and $\mathbf{Y}^{An}$, respectively. The following two self-learning rules are proposed to refine the interlaced instances in the training data set. (1) Any instance whose corresponding value in $\mathbf{R}_1^{An}$ is larger than the average value of $\mathbf{R}_1^{An}$ will be removed; (2) Any instance whose corresponding value in $\mathbf{R}_1^{Ae}$ is smaller than half of the average value of $\mathbf{R}_1^{Ae}$ will be removed.

C. *Rule-based Data Mining*

In the proposed framework, the C4.5 decision tree [15] is used for final event detection as it is a well-know rule-based algorithm good at learning the associations among different features of a set of pre-labeled instances. The filtered training data by the distance-based data mining scheme are fed into the C4.5 classifier to construct the tree model. Each data instance consists of the filtered features as well as the classification label, either "yes" for positive instances or "non" for negative instances. The filtered testing data are processed by the constructed tree model for final goal event detection.

IV. EXPERIMENTAL RESULTS

Soccer goal event detection is used as the experiment test bed to validate our proposed data mining framework. In our empirical study, 3 soccer videos were collected from different Internet broadcasters.

A. Experiment Setup

The video data (2 video) are randomly selected for training and the rest (1 video) are adopted for testing. 10 such groups are formed randomly for 10-fold cross-validation, and thus totally 10 decision models are constructed and tested with the corresponding testing data sets. In the experiments, the parameters for self-learning are set to $T\ 1=T\ 2=30$ and $K=50$ as discussed in Section 3.2.1 based on empirical studies.

B. Performance Evaluation

After 50 times random selection and comparison, the selected typical goal instances can recognize 100% goal instances and reject about 85% non-goal instances; while the selected typical non-goal instances can recognize about 80% non-goal instances and reject 100% goal instances.

The Recall and Precision for the goal event can be defined as:

$Recall = Identified/(Identified+Missed)$;

$Precision = Identified/(Identified+Mis-identified)$

Table 1. Cross validation results for goal detection

No	Goal	Ident	Missed	MisIdent.	PC(%)	PR(%)
1	13	13	0	8	100	61.9
2	13	11	2	1	84.6	91.7
3	13	10	3	0	76.9	100
4	13	12	1	1	92.3	92.3
5	13	10	3	1	76.9	90.9
AVG					86.14	87.36

Table 1 shows the goal event detection performance of our proposed data mining framework, where "RC", "PR", and "Ident." denote "Recall", "Precision", and "Identified". The "Missed" column indicates the number of goal instances that are misclassified as non-goal, and the "Mis-Ident." Column indicates the number of non-goal instances that are misclassified as goal instances.

The benefit of the proposed framework should be noted that the feature set has been reduced to almost one half, which brings operational benefits such as less storage requirement for multimedia database, less training time, less testing time, simplified tree model, and avoidance of "curse of dimensionality".

V. CONCLUSION

Video event detection is of great importance in video indexing, retrieval, and summarization. This paper proposed an advanced framework that utilizes both the distance-based and rule-based data mining techniques for domain independent video event detection to address the rare event detection issue. The proposed framework is fully automatic without the need of any domain knowledge, which is achieved by data pre-processing including increasing the percentage of positive instances and reducing the feature dimension, and a decision tree classifier for event detection. The experimental results in goal event detection from multiple broadcast video data show the viability and effectiveness of the proposed framework for general event detection.

REFERENCES

[1] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 796–807, Jul. 2003.

[2] B. Han, Support Vector Machines Center for Information Science and Technology, Temple University, Philadelphia, PA, 2003 [Online]. Available: http://www.ist.temple.edu/~vucetic/cis526fall2003/lecture8.doc

[3] D. Sadlier and N. E. O'Connor, "Event detection in field-sports video using audio-visual features and a support vector machine," *IEEE Trans. Circuits Syst. Video Technology*, vol. 15, no. 10, pp. 1225–1233, Oct. 2005.

[4] E.A. Tekalp and A.M. Tekalp, "Generic play-break event detection for summarization and hierarchical sports video analysis," *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 169–172, July 6-9, 2003.

[5] J. R. Quinlan*, C4.5: Programs for Machine Learning*. San Francisco, CA: Morgan Kaufmann, 1993.

[6] L. Xie, S.-F. Chang, A. Divakaran, and H. Sun, "Structure analysis of soccer video with hidden markov models," *Proc. IEEE IEEE International Conference on Acoustics, Speech, And Signal Processing (ICASSP),* vol. 4, pp. 4096–4099, May 13-17, 2002.

[7] M. Chen, S.-C. Chen, M.-L. Shyu, and K. Wickramaratna,"Semantic event detection via temporal analysis and multimodal data mining," *IEEE Signal Processing Magazine, Special Issue on Semantic Retrieval of Multimedia*,vol. 23, no. 2, pp. 38–46, March 2006.

[8] M. R. Naphade and J. R. Smith, "On the detection of semantic concepts at TRECVID," in *Proc. 12th ACM Int. Conf. Multimedia*, New York, 2004, pp. 660–667.

[9] S.C. Chen, M.-L. Shyu, C. Zhang, and M. Chen, "A multimodal data mining framework for soccer goal detection based on decision tree logic," *Int. J. Comput. Applic. Technol.*, vol. 27, no. 4, pp. 312–323, 2006.

[10] S. Dagtas and M. Abdel-Mottaleb, "Extraction of TV highlights using multimedia features," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing*, Cannes, France, 2001, pp. 91–96.

[11] T. Quirino, Z. Xie, M.-L. Shyu, S.-C. Chen, and L. Chang, "Collateral representative subspace projection modeling for supervised classification," *Proc. of the 18th IEEE International Conference on Tools with Artificial Intelligence*, pp. 98–105, November 13-15, 2006.

[12] V. S. Tseng, C.-J. Lee, and J.-H. Su, "Classify by representative or associations (CBROA): A hybrid approach for image classification," in *Proc. 6th Int. Workshop on Multimedia Data Mining: Mining Integrated Media and Complex Data*, Chicago, IL, Aug. 2005, pp. 37–53.