

Research Article

Video Genre Classification Using Weighted Kernel Logistic Regression

Ahmed A. M. Hamed,^{1,2} Renfa Li,¹ Zhang Xiaoming,¹ and Cheng Xu¹

¹ Hunan University, Changsha 410082, China

² University of Bahri, Khartoum 11123, Sudan

Correspondence should be addressed to Zhang Xiaoming; xiaoming.zhang@ymail.com

Received 28 March 2013; Accepted 10 July 2013

Academic Editor: Tai-hoon Kim

Copyright © 2013 Ahmed A. M. Hamed et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the widening semantic gap of videos, computational tools to classify these videos into different genre are highly needed to narrow it. Classifying videos accurately demands good representation of video data and an efficient and effective model to carry out the classification task. Kernel Logistic Regression (KLR), kernel version of logistic regression (LR), proves its efficiency as a classifier, which can naturally provide probabilities and extend to multiclass classification problems. In this paper, Weighted Kernel Logistic Regression (WKLR) algorithm is implemented for video genre classification to obtain significant accuracy, and it shows accurate and faster good results.

1. Introduction

Recently, video usage gains increasing importance, especially with the advance of recent internet technology and digital media; people have access to huge amount of data through internet and television. It is difficult for people to find videos of interest among these tremendous amounts of data and use them when there is a need, and it is not feasible to watch all the videos searching for the one of interest. Exploiting content-based of video footage is continuously needed for many applications, for example, for retrieving video sequence, creating automatic video summarization, or detecting specific action or activities in a video surveillance [1]. Many works are done dealing with video classification problem by categorizing videos in certain categories or genre, bridging the wide semantic gap between computed low-level features and high-level concepts and helping people to find their videos of interest within narrow domain. To get good understanding of video content, many different techniques have been developed and different video features have been identified for better video representation. Many techniques are used for video classification such as Bayesian, Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Neural-Network (NN), and Support Vector Machine (SVM).

SVM is considered a state-of-the-art algorithm for classifying binary data through its implementation of kernels. Oh et al. developed general framework to perform the fundamental tasks for video data mining which are temporal segmentation of video sequences and feature extraction and how to capture the location of motions occurring in a segment, how to cluster those segmented pieces, and how to find whether a segment has normal or abnormal events [2]. Chen et al. proposed a multimedia data mining framework for extraction of soccer goal events in soccer videos, by using combined multimodal analysis and decision tree [3]. Li et al. also proposed a general framework for indexing and summarizing sports broadcast programs, with a high-level model of sports broadcast video using the concept of an event, defined according to domain specific knowledge for different types of sports [4]. Zhong and Chang present a general framework and effective algorithms to detect the syntactic structures that are at a level higher than shots [5]. Fan et al. propose framework to achieve more effective semantic video classification by integrating the unlabeled samples with a limited number of labeled samples [6].

Petrushin describes an approach to efficient search for rare and frequent video events and builds a robust GMM

classifier to identify real time incoming events for indoor surveillance [7]. Hervieu et al. proposed a trajectory-based HMM framework for video content understanding to solve three tasks: unsupervised clustering of events, recognition of events corresponding to learned classes of dynamic video contents, and detection of unexpected events [1]. Adam et al. present a novel algorithm based on multiple local monitors which collect low-level statistics for detection of certain types of unusual events [8]. Shirahama et al. developed a method for retrieving events of interest in a video archive using “rough set theory” to extract multiple classification rules, each of which correctly identifies a subset of shots of the event [9].

Maalouf et al. [10, 11] show that Kernel Logistic Regression (KLR), kernel version of logistic regression (LR), proves its efficiency as a classifier, which can naturally provide probabilities and extend to multiclass classification problems. KLR is used for many reasons; the first is that LR and KLR are well studied, and they do not make assumption about the distribution of the independent variables, and last reason they can be extended to handle the multiclass classification problems. Thus with the right algorithm, the computation time can be reduced less than that for the other methods, such as Support Vector Machine (SVM), which require solving constrained quadratic optimization problem.

This paper is going to implement Weighted Kernel Logistic Regression (WKLR) algorithm for video genre classification, because it has direct probabilistic output and it could be extended directly to handle multiclass classification, not like SVM. The rest of the paper is organized as follows: Section 2 explains Logistic Regression and Kernel Logistic Regression model, Section 3 describes the Video Feature Extraction, Section 4 presents the proposed method, Section 5 experimental results are depicted, and Section 6 concludes the paper.

2. Logistic Regression and Kernel Logistic Regression Model

2.1. Logistic Regression. In logistic regression, a single outcome variable, y_i ($i = 1, \dots, n$), is coded 1 with probability p_i and 0 with probability $1 - p_i$. Then p_i varies as a function of some explanatory variables, such as x_i mathematically it is expressed as the following formula Maalouf et al. and King and Zeng [10, 12]:

$$E[y_i | x_i, \beta] = p_i = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}}, \quad (1)$$

where β is a vector of parameters with the assumption that $X_{j0} = 1$ so that the intercept β_0 is a constant term. So that the assumption that the intercept is included in the vector β will continue from now on. The logistic (logit) transformation is the logarithm of the odds of the positive response and is defined as

$$\eta_i = \ln\left(\frac{p_i}{1 - p_i}\right) = x_i \beta. \quad (2)$$

The logit function can be expressed in matrix form as

$$\eta = X\beta. \quad (3)$$

The regularized log likelihood is defined as

$$\ln L(\beta) = \sum_{i=1}^N (y_i \ln p_i + (1 - y_i) \ln(1 - p_i)) - \frac{\lambda}{2} \|\beta\|^2, \quad (4)$$

where the regularization (penalty) term $(\lambda/2)\|\beta\|^2$ is added to obtain better generalization.

For binary outputs, the loss function or the deviance DEV is the negative log likelihood and, given by the formula

$$\text{DEV}(\hat{\beta}) = -2 \ln L(\beta). \quad (5)$$

2.2. Kernel Logistic Regression. For Kernel logistic regression and following Maalouf et al., vector β can be expressed as a linear combination of the input vectors as the following [10]:

$$\beta = X^T \alpha = \sum_{i=1}^N \alpha_i X_i, \quad (6)$$

where the vector α is known as the dual variable with dimensions, $N \times 1$. Now the logit vector η can be rewritten as

$$\begin{aligned} \eta &= XX^T \alpha \\ &= K\alpha, \end{aligned} \quad (7)$$

where $K = XX^T$ is symmetric positive semidefinite Gram matrix with $N \times N$ dimensions.

Considering the logit link function again shown previously,

$$\eta_i = x_i \beta = \beta_0 + x_{i1} \beta_1, \dots, x_{id} \beta_d, \quad (8)$$

where the vector x_i is given by $x_i = [1, x_{i1}, \dots, x_{id}]$, with $i = 1, \dots, N$. this linear function represents the simplest form of an identity mapping polynomial basis function θ of the feature space such that

$$\theta(x_i) = \theta[(x_{i1}, \dots, x_{id})] = x_i. \quad (9)$$

So the logit link function can be rewritten as

$$\eta_i = \theta(x_i) \beta. \quad (10)$$

Generally the function $\theta(\cdot)$, maps the data from the lower dimension space to higher dimension space, so that

$$\theta : x \in R^d \longrightarrow \theta(x) \in F \subseteq R^s. \quad (11)$$

The purpose of choosing the mapping θ is to convert nonlinear relation between the response variable and the independent variables into linear relation, Maalouf and Trafalis [11].

Logit link function in case of KLR then could be expressed as

$$\begin{aligned}\eta_i &= \sum_{j=1}^N \alpha_j \langle \theta(x_i), \theta(x_j) \rangle \\ &= \sum_{j=1}^N \alpha_j k(x_i, x_j) \\ &= k_i \alpha,\end{aligned}\quad (12)$$

where k_i is the i th row in the kernel matrix $k(x_i, x_j) = K$. Mercer's sufficient and necessary conditions must be satisfied in kernel transformation function, which state that the kernel function must be expressed as inner product and must be positive and semidefinite, Maalouf and Trafalis [11]. Now

$$\eta_i = k_i \alpha = \ln \left(\frac{p_i}{1 - p_i} \right). \quad (13)$$

This implies that

$$p_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}} = \frac{e^{k_i \alpha}}{1 + e^{k_i \alpha}} = \frac{1}{1 + e^{-k_i \alpha}}. \quad (14)$$

So the regularization likelihood can be rewritten with respect to α as

$$\ln L(\alpha) = \sum_{i=1}^N (y_i \ln p_i + (1 - y_i) \ln(1 - p_i)) - \frac{\lambda}{2} \alpha^T K \alpha. \quad (15)$$

With deviance

$$\text{DEV}(\alpha) = -2 \ln L(\alpha), \quad (16)$$

iteratively reweighted least squares (IRLS) method is one of the most popular techniques used to find the MLE of β in LR models, which uses Newton-Raphson algorithm to solve the score equations. KLR models can also be fitted using IRLS. Each iteration finds the weighted least squares (WLS) estimates for a given set of weights, which are used to construct a new set of weights Oh et al. [2]. The gradient and Hessian are obtained by differentiating $\ln L(\alpha)$ with respect to α . In matrix form, the gradient is

$$\nabla_{\alpha} \ln L(\alpha) = K^T (y - p) - \lambda K \alpha = 0, \quad (17)$$

where p is the probability vector whose elements are given in (14). And the Hessian with respect to α is

$$\nabla_{\alpha}^2 \ln L(\alpha) = -K^T V K - \lambda K, \quad (18)$$

where V is a diagonal matrix with diagonal elements $p_i(1 - p_i)$ for $i = 1, \dots, N$. The Newton-Raphson update with respect to α on the $(c + 1)$ -th iteration is

$$\hat{\alpha}^{(c+1)} = \hat{\alpha}^{(c)} + (K^T V K + \lambda K)^{-1} K^T (y - p) - \lambda K \alpha^{(c)}. \quad (19)$$

Since $\hat{\alpha}^{(c)} = (K^T V K + \lambda K)^{-1} (K^T V K + \lambda K) \hat{\alpha}^{(c)}$, then (10) can be rewritten as

$$\hat{\alpha}^{(c+1)} = (K^T V K + \lambda K)^{-1} K^T V z^{(c)}, \quad (20)$$

where $z^{(c)} = K \hat{\alpha}^{(c)} + V^{-1}(y - p)$ is the adjusted dependent variable or adjusted response.

2.3. *WKLR Algorithm.* Consider

$$(K^T V K + \lambda K) \hat{\alpha}^{(c+1)} = K^T V z^{(c)} \quad (21)$$

is a system linear equations with kernel matrix K , vector of adjusted responses z , and weighted matrix V , and the weights and response vector both depend on $\hat{\alpha}^{(c)}$, which is the current estimate of the parameter vector, so specifying an initial estimate $\hat{\alpha}^{(0)}$ for $\hat{\alpha}$ can be solved iteratively, using the Conjugate Gradient (CG) method, which is equivalent to minimizing the quadratic problem

$$\frac{1}{2} \hat{\alpha}^{(c+1)} (K^T V K + \lambda K) \hat{\alpha}^{(c+1)} - \hat{\alpha}^{(c+1)} (K^T V z^{(c)}). \quad (22)$$

In order to avoid the long computations that the CG may suffer from, a limit can be placed on the number of CG iterations, thus creating an approximate or truncated Newton direction, Maalouf and Trafalis [11].

Algorithm 1 describes Iterative Weighted Kernel Logistic Regression Maximum Likelihood Estimate (IWKLR MLE) using IRLS.

3. Video Feature Extraction

Extracting appropriate features is vital for acceptable design of any pattern classifier. In video classification studies, the features used can be categorized as one of three which are text features, audio features, and visual features; some studies divide the features into two main categories: text features and nontext features, and the latter is divided into low-level and semantic video features. Text features contain text extracted from video and user-generated text feature. Nontext features are features extracted from images, audio, and motion; such features could be referred to as low-level features which could be defined as features extracted from video clips and audio track without reference to any external knowledge [13, 14]. Some of researches use features that correspond to cinematic principle for visual features which are composed of color, motion and average shot length. One difficulty in using low-level feature especially visual features is the huge amount of data, and the most used solution for this is using a key frame to represent the shot or by using dimensionality reduction techniques such as Principal Component Analysis (PCA) and wavelet transforms application [13, 15, 16]. Discrete transform is widely used for feature extraction and data redundancy reduction. Proportional advantages of deterministic transforms make them an interesting type of feature extraction approaches. One of the important discrete transform, is Discrete Cosine Transform (DCT), and special properties of the DCT make it a powerful transform in video processing applications [17–20]. DCT has ability for data decorrelation. There are fast algorithms for DCT realization. When applying DCT to video, some coefficients are selected and others discarded for dimension reduction of data. DCT coefficient selection is an important part of feature extraction process. DCT feature extraction is composed of two steps: the first one DCT is applied to the entire key frame to obtain the DCT coefficients, then in the second step some of coefficients

```

Data:  $K, y, \hat{\alpha}^{(0)}$ 
Result:  $\hat{\alpha}$ 
Begin
   $C = 0$ 
  while  $\left| \frac{\text{DEV}(c) - \text{DEV}(c+1)}{\text{DEV}(c+1)} \right| > \varepsilon_1$  and  $C \leq \text{MAX ILRS Iteration}$  Do
    for  $i \leftarrow 1$  to  $n$  do
       $\hat{p}_i = \frac{1}{1 + e^{-x_i \hat{\alpha}}}$ ; // compute probabilities
       $v_i = \hat{p}_i(1 - \hat{p}_i)$ ; // compute weights
       $z_i = K_i \hat{\alpha}^{(c)} + \frac{(y_i - \hat{p}_i)}{\hat{p}_i(1 - \hat{p}_i)}$ ; // compute adjusted response
     $V = \text{diag}(v_1, \dots, v_N)$ ;
     $(K^T V K + \lambda K) \hat{\alpha}^{(c+1)} = K^T V z^{(c)}$ ; // compute  $\hat{\alpha}$ 
     $C = C + 1$ 
End

```

ALGORITHM 1

are selected to construct feature vectors. The size of the DCT coefficient matrix is the same as the input frame. DCT by itself does not decrease data dimension, so it compresses most signal information in small percent of coefficients. DCT coefficients for $M \times N$ frame are calculated as follows [18]:

$$\begin{aligned}
f(u, v) &= \frac{1}{\sqrt{MN}} \alpha(u) \alpha(v) \\
&\times \sum_{x=0}^{M-1-N} \sum_{y=0}^{N-1} f(x, y) \times \cos\left(\frac{(2x+1)u\pi}{2M}\right) \times \cos\left(\frac{(2y+1)v\pi}{2N}\right), \\
&u = 0, 1, \dots, M, v = 0, 1, \dots, N,
\end{aligned} \tag{23}$$

where $\alpha(\omega)$ is defined by

$$\alpha(\omega) = \begin{cases} 1 & \omega = 0. \\ \sqrt{2} & \omega \neq 0. \end{cases} \tag{24}$$

$f(x, y)$ is the frame intensity function, and $F(u, v)$ is 2D matrix of DCT coefficients. As mentioned that DCT coefficient matrix is the same as the frame size and to obtain feasible and compact low-dimensional representation of the features, PCA could be used [21]. Given N d -dimensional feature vectors [22, ..., N], the mean vector calculated as

$$\tau = \frac{1}{N} \sum_{i=1}^N x_i. \tag{25}$$

And covariance matrix is calculated as

$$c = \frac{1}{N} \sum_{i=1}^N (x_i - \tau)(x_i - \tau)^T; \tag{26}$$

then the Principal Component assumes the first P significant eigenvectors of c , that is, $\{v_i, i = 1, 2, \dots, P\}$. By constructing

the eigenmatrix $U = [v_1, v_2, \dots, v_P]$ of $d \times P$ dimension, an arbitrary d -dimensional original feature vector x can be represented as a new low P -dimensional vector $y = U^T(x - \tau)$, and $P \ll N$, and $P \ll d$ [21]. In our proposed method we deal with shot or scene concept and take a key frame to represent the whole shot or scene. There are different algorithms implemented for extracting key frame; most of these algorithms depend on the motion part to extract the key frames. After extracting the key frames, we used DCT to obtain the DCT coefficients. PCA is applied on the DCT coefficients to select the most significant DCT coefficients.

4. Proposed Method

In this paper, first shot detection is used to extract the key frames from the input videos; shot detection is used because it is more sophisticated method for video summary [19, 22]. Secondly we construct X and Y variables; variable represents the features of video shot represented by the DCT_PCA data, and Y variable determines that video shot is set manually. Then feature should be scaled to values between 0 and 1 using the following equation:

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{27}$$

The advantage of this scaling is to avoid features in greater numeric range dominating those in smaller numeric range. Another advantage is to avoid numerical difficulties during the calculation when large feature value might cause numerical problem. Finally WKLR is implemented on the prepared data for classification aiming to achieve significant accuracy and making WKLR to be an effective method for video classification.

4.1. Performance Measures. The predictions quality evaluated using five measures: Matthews correlation coefficient (MCC), Q_{total} , Predicted Positive Value (PPV), sensitivity, and specificity. FP: false positive, FN: false negative, TP: true positive,

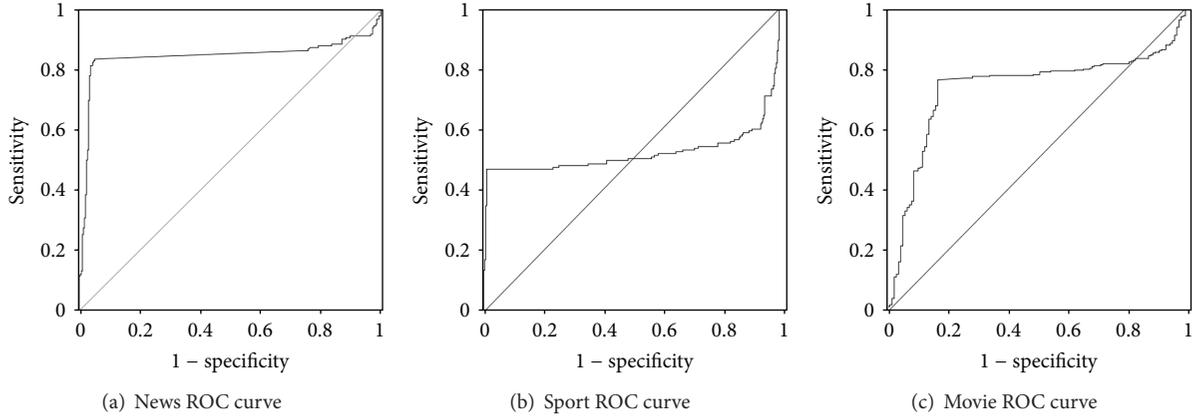


FIGURE 1: ROC curve for news, sport, and movie, respectively.

TABLE 1: Number of positive and negative occurrences for each genre.

	TP	TN	FP	FN
News	829	1322	119	109
Sport	664	1402	129	179
Movies	449	1621	134	169

and TN: true negative. Matthews correlation coefficient can be in the range of -1 to 1 , where 1 is a perfect correlation and -1 is the perfect anticorrelation. A value of 0 indicates no correlation. Its value can be calculated as follows:

$$\begin{aligned} \text{MCC} \\ = \frac{(\text{TP} * \text{TN} - \text{FP} * \text{FN})}{\sqrt{(\text{TP} + \text{FN}) * (\text{TN} + \text{FP}) * (\text{TP} + \text{FP}) * (\text{TN} + \text{FN})}}. \end{aligned} \quad (28)$$

Q_{total} is the percentage of correctly classified residues, also called the prediction accuracy. It is given as follows:

$$Q_{\text{total}} = \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}. \quad (29)$$

PPV is the predicted positive value, also called the precision or Q_{pred} . It is given as follows:

$$Q_{\text{pred}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100. \quad (30)$$

Sensitivity is also called recall or Q_{obs} and is the fraction of the total positive examples that are correctly predicted, and it is calculated as follows:

$$Q_{\text{obs}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100. \quad (31)$$

Specificity is the fraction of total negative examples that are correctly predicted; its value can be calculated as follows:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100. \quad (32)$$

5. Experimental Results

The steps mentioned previously are tested on group of videos downloaded from YouTube and Youku websites, containing the three categories, News, Sport, and Movies, and the selected frames of all those videos are almost about 10000 frames to test the accuracy of the proposed method. In the training phase, the data distributed randomly, the algorithm of WKLR MLE was implemented with 10-fold cross-validation, and max number of iterations is set to 30 for truncated Newton method.

Figure 1, Tables 1 and 2 show the accuracy and classification performance of each category against the other two categories.

As shown in Table 2, using KLR with same Gaussian kernel shows that the proposed approach could achieve considerable improvement in performance on small to medium size data sets; the average accuracy of our proposed method is 88%, compared to 86% accuracy of the work done in [23] in which the last paper proposed a new technique for improving unconstrained face recognition based on leveraging weakly labeled web videos. KLR has also advantages including that KLR has direct extension to multiclass classification and it can yield probabilistic output without intervention.

6. Conclusion

This paper proposed Weighted Kernel Logistic Regression (WKLR) in order to enhance the accuracy and processing time of video classification, which shows easy to implement and good acceptable results regarding to performance. One of the benefits of using WKLR is that it uses algorithms which are less complex for unconstrained optimization methods compared to algorithms used for constrained optimization

TABLE 2: The accuracy and MCC for each genre.

	Accuracy	MCC	Qpred	Qobs	Specificity
News	90%	80%	88%	88%	92%
Sport	87%	71%	84%	79%	92%
Movies	87%	66%	77%	72%	93%

methods such as SVM. An other benefit is that WKLR takes the advantage of speed of the IRLS and the power of the kernel methods. Future studies may compare this method using different kernels and test the algorithm on more datasets. They may also try to improve the speed of the algorithm. Methods such as the trust region Newton might also be utilized because of their stability and rigorous mathematical derivation.

References

- [1] A. Hervieu, P. Boutheymy, and J. P. L. Cadre, "Video event classification and detection using 2D trajectories," in *Proceedings of the 3rd International Conference on Computer Vision Theory and Applications (VISAPP '08)*, pp. 158–166, January 2008.
- [2] J. H. Oh, J. Lee, and S. Kote, "Multimedia data mining framework for raw video sequences," in *Mining Multimedia and Complex Data*, pp. 18–35, 2003.
- [3] S. C. Chen, M.-L. Shyu, M. Chen, and C. Zhang, "A decision tree-based multimodal data mining framework for soccer goal detection," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, pp. 265–268, IEEE, June 2004.
- [4] B. Li, J. H. Errico, H. Pan, and I. Sezan, "Bridging the semantic gap in sports video retrieval and summarization," *Journal of Visual Communication and Image Representation*, vol. 15, no. 3, pp. 393–424, 2004.
- [5] D. Zhong and S. F. Chang, "Real-time view recognition and event detection for sports video," *Journal of Visual Communication and Image Representation*, vol. 15, no. 3, pp. 330–347, 2004.
- [6] J. Fan, H. Luo, J. Xiao, and L. Wu, "Semantic video classification and feature subset selection under context and concept uncertainty," in *Proceedings of the 4th ACM/IEEE Joint Conference on Digital Libraries: Global Reach and Diverse Impact (JCDL '04)*, pp. 192–201, IEEE, June 2004.
- [7] V. A. Petrushin, "Mining rare and frequent events in multi-camera surveillance video using self-organizing maps," in *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '05)*, pp. 794–800, ACM, August 2005.
- [8] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, 2008.
- [9] K. Shirahama, Y. Matsuoka, and K. Uehara, "Video event retrieval from a small number of examples using rough set theory," in *Advances in Multimedia Modeling*, pp. 96–106, 2011.
- [10] M. Maalouf, T. B. Trafalis, and I. Adrianto, "Kernel logistic regression using truncated Newton method," *Computational Management Science*, vol. 8, no. 4, pp. 415–428, 2011.
- [11] M. Maalouf and T. B. Trafalis, "Robust weighted kernel logistic regression in imbalanced and rare events data," *Computational Statistics & Data Analysis*, vol. 55, no. 1, pp. 168–183, 2011.
- [12] G. King and L. C. Zeng, "Explaining rare events in international relations," *International Organization*, vol. 55, no. 3, pp. 693–715, 2001.
- [13] D. Brezeale and D. J. Cook, "Automatic video classification: a survey of the literature," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 38, no. 3, pp. 416–430, 2008.
- [14] A. A. M. Hamed, Z. Xiaoming, C. Xu et al., "Video genre classification using support vector machine ensemble," *International Journal of Digital Content Technology and Its Applications*, vol. 6, no. 15, pp. 191–200, 2012.
- [15] H. Zhou, T. Hermans, A. V. Karandikar, and J. M. Rehg, "Movie genre classification via scene categorization," in *Proceedings of the International Conference on Multimedia (MM '10)*, pp. 747–750, ACM, October 2010.
- [16] Z. Rasheed, Y. Sheikh, and M. Shah, "On the use of computable features for film classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 52–64, 2005.
- [17] A. Girgensohn and J. Foote, "Video classification using transform coefficients," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)*, pp. 3045–3048, IEEE, March 1999.
- [18] S. Dabbaghchian, M. P. Ghaemmaghami, and A. Aghagolzadeh, "Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology," *Pattern Recognition*, vol. 43, no. 4, pp. 1431–1440, 2010.
- [19] C. A. Dhawale and S. Jain, "A novel approach towards keyframe selection for video summarization," *Asian Journal of Information Technology*, vol. 7, no. 4, pp. 133–137, 2008.
- [20] M. Heckmann, K. Kroschel, C. Savariaux et al., "DCT-based video features for audio-visual speech recognition, 2002."
- [21] L. Q. Xu and Y. Li, "Video classification using spatial-temporal features and PCA," in *Proceedings of the International Conference on Multimedia and Expo (ICME '03)*, vol. 3, pp. 485–488, IEEE, July 2003.
- [22] M. Mentzelopoulos and A. Psarrou, "Key-frame extraction algorithm using entropy difference," in *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR '04)*, pp. 39–45, ACM, October 2004.
- [23] D. Rim, K. Hassan, and C. Pal, "Semi supervised learning in wild faces and videos," in *Proceedings of the British Machine Vision Conference*, J. Hoey, S. McKenna, and E. Trucco, Eds., pp. 3.1–3.12, September 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

