Fast Film Genres Classification Combining Poster and Synopsis

Zhikang Fu[1], Bing Li[2], Jun Li[3], and Shuhua Wei[1](✉)

[1] College of Electronic Information Engineering,
North China University of Technology, Beijing, China
fzklove@126.com, jslwsh@hotmail.com

[2] Chinese Academy of Sciences, Institute of Automation, Beijing, China
bli@nlpr.ia.ac.cn

[3] School of Automation, Southeast University, Nanjing, China
Lijun_automation@seu.edu.cn

Abstract. In this paper, we present an efficient approach to fast classify film genre by making use of film posters and synopsis simultaneously. Compared with traditional video content-based classification methods, the proposed method is much faster and more accurate. In the proposed method, a film poster is represented as multiple features including color, edge, texture, and the number of faces. On the other hand, we employ Vector Space Model (VSM) to characterize the texts in the synopsis. Then, we train a poster classifier and a text classifier using the Support Vector Machine (SVM). Finally, a test film is classified based on the 'OR' operation on the outputs of the two classifiers. We verify our scheme on our collected film poster and synopsis dataset. The experimental results demonstrate the promise of our method which achieves the desirable performance by combining posters with synopsis.

Keywords: Film genre · Film poster · VSM · Synopsis · SVM · OR

1 Introduction

More and more films come into our life along with the rapid development of the Internet. Recent years witness the extensive research conducted in the film genre classification. However, limited progress has been made due to the challenge of big data and the ambiguity in the definition of film genres. In this paper, we classify the film into four categories as illustrated in Fig. 1, and present a generic framework for film genre classification.

1.1 Related Work

Rasheed et al. [1] extracted low-level visual features from movies manually and classified them into four genres, namely: drama, action, comedy, horror. Zhou et al. [2] simultaneously adopted three kinds of features, i.e. GIST, CENTRIST and W-CENTRIST scene features, to describe a collection of temporally-ordered static key frames for the sake of representation. Genre classification and test on 1239 movies

X. He et al. (Eds.): IScIDE 2015, Part I, LNCS 9242, pp. 72–81, 2015.
DOI: 10.1007/978-3-319-23989-7_8

trailers were based on visual vocabulary structured by these features. Huang et al. [3] employed the same features used in [1] to categorize movies into three genres which are action, drama, and thriller. Ivasic-Kos et al. [4] utilized film posters to achieve effective film genre classification. Specifically, they proposed to use a set of low-level features for multi-label poster classification. The multi-label poster classification refers to the scenario in which the film poster simultaneously contains two informative labels from the label set of action, animation, comedy, drama, horror and war, which poses more challenges in contrast to the conventional genre classification problem in which only a single label is taken into account. But, its accuracy is very low for inadequate features. Subashini et al. [11] proposed a method for combing audio and video for classifying the genre of a movie. His results are better, but vast audio data and video data are used for his experiments. Paris et al. [12] made use of a thematic intensity extracted from synopsis and movie content to detect animated movies. However, his method can not detect other genres of films.

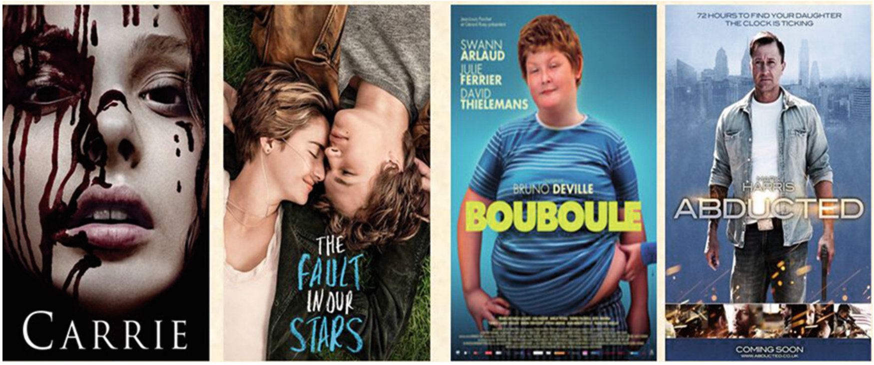

Fig. 1. The illustrative four films genres (from left to right are horror, love, comedy and action)

1.2 Our Work

On account of different definitions of film genres currently, it is required to determine the task-specific film genres beforehand. Specifically, the classical genres available on the popular film websites are given in Table 1. Without loss of generality, we refer to massive relevant literatures [1, 3, 4] and divide films into four groups: the horror films, comedies, love stories and action movies.

In this paper, we take advantage of film posters and synopsis to classify films into four genres. We train image set and text set by SVM separately. So, we get two respective predictions. If any prediction is right, we choose the right prediction as the last prediction of a film. Otherwise the last prediction is decided by the prediction based on poster.

The rest of the paper is organized as follows: Sect. 2 gives the whole framework of our proposed method. The following Sect. 3 introduces our elaborated devised features extracted from images and texts. We provide our experimental results in Sect. 4. Section 5 concludes this paper.

Table 1. The illustrative film genres on major domestic and foreign film website

	Genres
So Hu	Comedy Love Action Thriller War Science Fiction Disaster
Teng Xun	Action Adventure Comedy Love War Crime Thriller Science Fiction
You Ku	Comedy Horror Love Action Science Fiction War Crime
iqiyi	Love Comedy Action Horror Ethic Science Fiction Crime
YouTube	Adventure Animation Comedy Drama horror Love Action
The Movie DataBase	Thriller Adventure Science Fiction Romance Action Crime Horror Drama

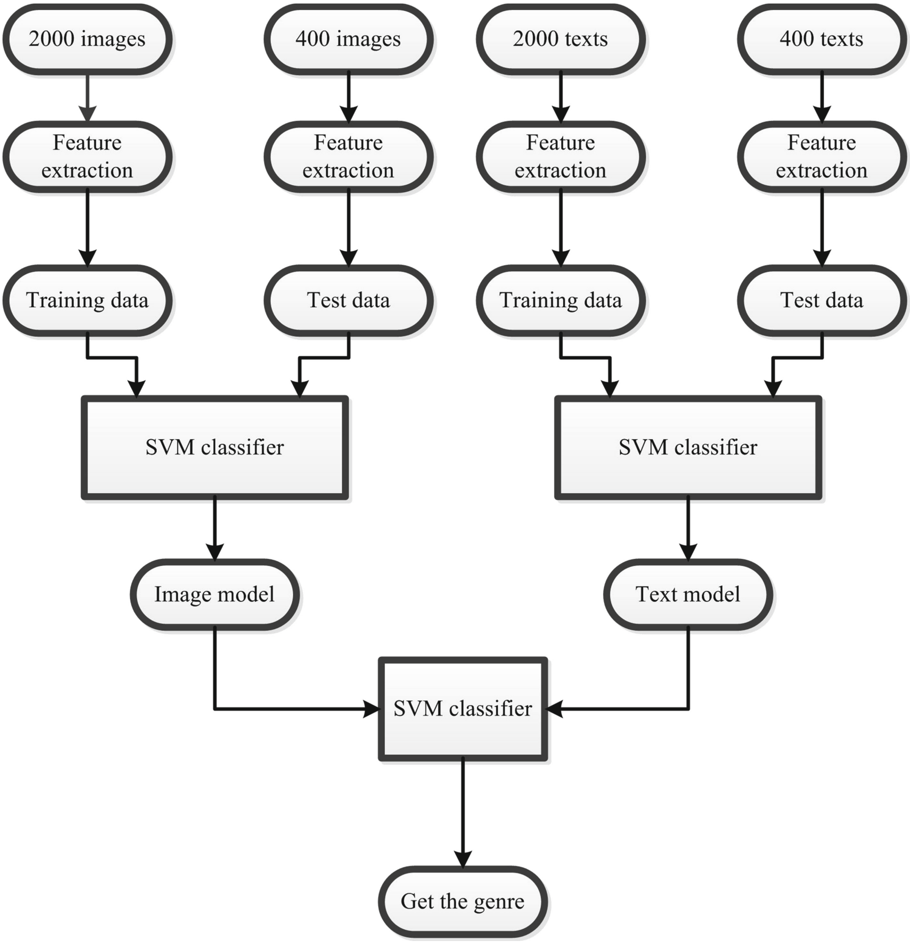

Fig. 2. The frame work of our method

2 Proposed Method

Figure 2 illustrates the processing pipeline of our method. First, we obtain high-resolution film posters and relevant synopsis from several popular foreign film websites. We operate on six feature modalities for the fine description of posters: color emotion, color harmony, edge feature, texture, color variance and the number of faces. Besides, the film synopsis is represented by VSM. We put them into SVM classifier separately and derive respective detectors from dual modalities, namely image and text model. Then, we get the first prediction Y1based on image and the second predictionY2 based on text. If any prediction is true, we choose the right prediction as the last prediction Y of a film. Otherwise Y is decided by theY1.

We simultaneously employ film posters and synopsis to detect film genres for many advantages.

- "Fast." It is faster to get the detector of a film, comparing with using the video content.
- "Accuracy." We get a high accuracy with combing posters and synopsis. The last result is up to 88.5 %.
- "Convenience." We can get the genre of a film with its poster and synopsis at the situation of non-existent video content.

Last, we classify films by posters or synopsis singly, comparing with our method.

3 Feature Extraction

Under our framework, image features and text features are simultaneously extracted. Specifically, the features of film posters are obtained by utilizing six low-level attributes: color emotion, color harmony, edge feature, texture, color variance and the number of faces. Additionally, the texts in the film synopsis are described by making use of VSM. The feature generations are detailed in the following sections.

3.1 Image Feature

Color Emotion. In real world, color is a chromatic cue which significantly influences our emotion and feelings. We respond to different colors in very different moods. For example, we are likely to feel excited, nervous while in an environment full of red objects. Conversely, lush scenery can make us feel light-hearted and comfortable. Likewise, blue enables bringing us the feeling of warmness and serenity.

In order to better delineate the color and correlate it with human emotion in mathematical formulation, Ou et al. [5, 6] proposed that human emotion are closely related with three factors relevant to color cues: Activity, Weight, and heat:

$$\begin{aligned} activity &= -2.1 + 0.06\left[(a^{*}-3)^{2} + (L^{*}-50)^{2} + (\frac{b^{*}-17}{1.4})^{2}\right]^{1/2} \\ weight &= -1.8 + 0.45\cos(h-10^{o}) + 0.04(100-L^{*}) \\ heat &= -0.5 + 0.02(C^{*})^{1.07}\cos(h-50^{o}) \end{aligned} \tag{1}$$

Where $(L*, C*, h)$ and $(L*, a^{*}, b^{*})$ are the color values in CIELCH and CIELAB color spaces respectively.

We define each pixel's color emotion EI as:

$$EI(x,y) = \sqrt{activity^{2} + weight^{2} + heat^{2}.} \tag{2}$$

Color Harmony. Color harmony of two-color combinations has been investigated in several empirical experiments. Ou et al. [7] proposed a model based on a psychophysical experiment of two-color combinations for predicting color harmony of two-color combinations. The model includes H_H (hue effect), H_L lightness effect and H_C (chromatic effect.)

$$\begin{aligned} H_C &= 0.04 + 0.53\tanh(0.8 - 0.045\Delta C) \\ \Delta C &= \left[(\Delta H^{*}_{ab})^{2} + (\frac{\Delta C^{*}_{ab}}{1.46})^{2}\right]^{\frac{1}{2}} \\ H_L &= H_{Lsum} + H_{\Delta L} \\ H_{Lsum} &= 0.28 + 0.54\tanh(-3.88 + 0.029\Delta L_{sum}) \\ L_{sum} &= L^{*}_{1} + L^{*}_{2} \\ H_{\Delta L} &= 0.14 + 0.15\tanh(-2 + 0.2\Delta L). \\ \Delta L &= \left|L^{*}_{1} - L*_{2}\right|. \\ H_H &= H_{SY1} + H_{SY2} \\ H_{SY} &= E_C(H_S + E_Y) \\ E_C &= 0.5 + 0.5\tanh(-2 + 0.5C^{*}_{ab}) \\ H_s &= 0.08 - 0.14\sin(h_{ab} + 50^{o}) - 0.07\sin(2h_{ab} + 90^{o}) \\ E_Y &= \frac{0.22L^{*} - 12.8}{10}\exp\{\frac{90^{o} - h_{ab}}{10} - \exp\{\frac{90^{o} - h_{ab}}{10}\}\} \end{aligned} \tag{3}$$

Where h_{ab} = CIELAB hue angle, C^{*}_{ab} = CIELAB chroma, ΔC^{*}_{ab} and ΔH^{*}_{ab} are the difference of two-color in CIELAB color space, L^{*}_{1} and L^{*}_{2} are the lightness of two different colors in CIELAB color space. Color harmony (CH) is defined as:

$$CH = H_H + H_C + H_L. \tag{4}$$

Edge Feature. Given an image, we begin with its transform from RGB into HSV color space. The derived value (V) channel is blurred by the 3 × 3 Gaussian filter. Next, the result is convolved by the Sobel edge detector. Finally, the outlier pixels are filtered by using the predefined threshold which is empirically set to be 0.5 in our experiment.

Texture Feature. Geusebroek et al. [8] proposed a six-stimulus basis to express stochastic texture perception. The texture statistics of an image is assumed to drawn from Weibull-distribution.

$$wb(y) = \frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\frac{1}{\gamma}\left(\frac{y}{\beta}\right)^{\gamma}} Z \tag{5}$$

The parameters of the distribution enable the fine description of the spatial structure of the texture. The wild size is given by β which represents the contrast of an image while the gain size γ denotes the peakedness of the distribution.

Color Variance. To detect the color variability exhibited in the film poster, we employ the CIE*L*uv color space, since it is designed to match with human perception. The three-order covariance matrix ρ is defined as:

$$\rho = \begin{pmatrix} \sigma_L^2 & \sigma_{Lu}^2 & \sigma_{Lv}^2 \\ \sigma_{Lu}^2 & \sigma_u^2 & \sigma_{uv}^2 \\ \sigma_{Lv}^2 & \sigma_{uv}^2 & \sigma_v^2 \end{pmatrix}. \tag{6}$$

Color variance is thus represented by the determinant Δ_F:

$$\Delta_F = \det(\rho). \tag{7}$$

The Number of Faces. Our observation implies the absence of normal human faces in the horror film posters and frequent occurrences of frontal faces and profiles in the comedy posters. Thus, we consider the number of faces in the film poster as an independent feature and detect human faces in the poster. In implementation, the detection of front faces is achieved by employing OpenCV containing a haarcascade_frontalface_alt model. The illustrative result is demonstrated in Fig. 3.

3.2 Text Feature

The English film synopsis is crawled from the Movie Data Base (TMDB) [10] website. We adopt the BOWs framework for obtaining text feature. The synopsis of every film is taken as a text document, removing the stop word of every text document, getting the stem of every word in the text document with porter's [9] algorithm, selecting feature word with information gain, structuring the Bag-of-words based feature word and representing every text document in term of VSM.

Reduction of English Stem. There are many forms in the same English word, such as adjective tense, past tense, progressive tense and so on. So, we must get the stem of every word for reducing the dimension of features. It has been demonstrated that we can get a better result compare with others, using porter's algorithm for reducing English stem.

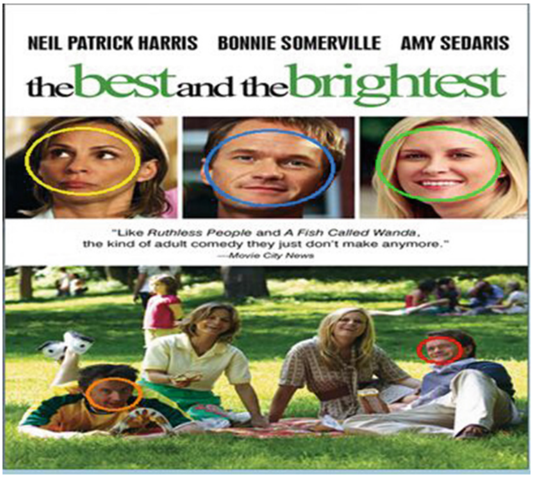
NEIL PATRICK HARRIS
BONNIE SOMERVILLE
AMY SEDARIS
the best and the brightest
"Like Ruthless People and A Fish Called Wanda, the kind of adult comedy they just don't make anymore."
—Movie City News

Fig. 3. The results of face detection

Structure of Bag-of-Words. We should have typical feature word which can represent the content of every document and the genres of films. Information Gain (IG) is used to choose the feature word in this paper. IG formula is following:

$$
\begin{aligned}
IG(T) &= H(C) - H(C|T) \\
H(C) &= -\sum_{i=1}^{n} P(C_i)\log_2 P(C_i). \\
H(C|T) &= -P(t)\sum_{i=1}^{n} P(C_i|t)\log_2 P(C_i|t) - P(\bar{t})\sum_{i=1}^{n} P(C_i|\bar{t})\log_2 P(C_i|\bar{t})
\end{aligned}
\tag{8}
$$

Where $P(t)$ is the document frequency of the feature word T. It counts the number of documents in S where T appears. |S| is the total number of documents in the corpus. $P(C_i|t)$ is the document frequency of the feature word T under the situation of C_i category. It counts the number of documents in D where T appears. |D| is the total number of documents in the C_i category.

Last Bag-of-Words is been constructed by feature word. At the same time every document can be described by VSM based on Bag-of-Words.

3.3 Classification

We construct two irrelevantly training set which are image training and text training. Then they are passed into SVM classifier. We get image model and text model. Subsequently, we get result Y1 by using image model to predict image test set. We get result Y2 by using text model to predict text test set. Last, if any prediction is true, we choose the right prediction as the last prediction Y of a film. Otherwise Y is decided by the Y1.

4 Experiments

4.1 Dataset

For performance measure of our proposed method, the experiments are carried out on the collection of websites including the English text and images. We collect 2400 film posters and 2400 text documents obtained from TMDB and select 4 genres (Horror, Comedy, Romance, Action) each of which has 600 training examples. We employ 2000 posters and 2000 text document for training and the rest are used for test. Our dataset is balanced. Each genre has 500 samples in training set. Meantime, each genre has 100 samples in test set.

4.2 The Result of Experiments

To demonstrate our proposed method, we have performed three experiments: film genres classification using posters, film genres classification using synopsis, and film genres classification combining posters and synopsis.

First, we extract the poster features, using SVM classifier which is double Radial Basis Function (RBF) kernel to predict the genre of a film. The result is shown in Table 2. We can see that there is low accuracy for classifying films by posters singly, especially the accuracy of action.

Table 2. The result of predicting the posters

	Horror	Comedy	Love	Action	All test set
Accuracy	67 %	61 %	64 %	51 %	60.75 %

Then, we extract the synopsis features. SVM classifier with RBF kernel is used for predict the genre of a film. The result is shown in Table 3. We can see that text features can perform better than image features. The accuracy of each genre has been improved obviously, especially the accuracy of action which is up to 89 %. Meantime, the computing time in second experiment is shorter than the previous experiment. However, the accuracy of comedy is very low according to Tables 2 and 3.

Table 3. The result of predicting the texts

	Horror	Comedy	Love	Action	All test set
Accuracy	70 %	62 %	72 %	89 %	73.25 %

Last, we feed image model and text model which are got from previous two experiments into the same SVM as before. We fuse the prediction further. The best result is shown in Table 4. We can see that the accuracy has exceeded 90 % in horror, love and action. The lowest is comedy which is 81 %. The accuracy of test set is up to 88.5 %. However, the computing time in third experiment is longer than the previous

Table 4. The result of predicting the fusion

	Horror	Comedy	Love	Action	All test set
Accuracy	91 %	81 %	90 %	92 %	88.5 %

two experiments. We come to the conclusion that classifying films combing posters and synopsis can get high accuracy.

5 Conclusions

In this paper, the genres of films are detected by combining posters and synopsis. The posters are detected with color emotion, color harmony, edge feature, texture, color variance and the number of faces. At the same time, the synopsis is represented in VSM. We employ image model to predict image test set and text model to predict text test set separately. The last fusion is based on the OR operation of two detectors. Experimental results show that the proposed method is fast, high accuracy and convenient in film classification.

Acknowledgements. This work is partly supported by the 973 basic research program of China (Grant No. 2014CB349303), the Natural Science Foundation of China (Grant No. 61472421), the National 863 High-Tech R&D Program of China (Grant No. 2012AA012504), and the Project Supported by Guangdong Natural Science Foundation (Grant No. S2012020011081), and the Scientic Research Project of Beijing Educational Committee (No. KM201410009005).

References

1. Rasheed, Z., Sheikh, Y., Shah, M.: On the use of computable features for film classification. IEEE Trans. Circ. Syst. Video Technol. **15**(1), 52–64 (2005)
2. Zhou, H., Hermans, T., Karandikar, A.V., Rehg, J.M.: Movie genre classification via scene categorization. In: International Conference on Multimedia, pp. 747–750 (2010)
3. Huang, H.-Y., Shih, W.-S., Hsu, W.-H.: A film classifier based on low-level visual features. In: International Workshop on Multimedia Signal Processing, pp. 465–468 (2007)
4. Ivasic-Kos, M., Pobar, M., Mikec, L.: Movie Posters Classification into Genres Based on Low-level Features. In: International Convention on Information and Communication Technology, pp. 1198–1203 (2014)
5. Ou, L.C., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. part I: colour emotions for single colours. Color Res. Appl. **29**(3), 232–240 (2004)
6. Ou, L.C., Luo, M.R., Woodcock, A., Wright, A.: A study of colour emotion and colour preference. part III: colour preference modeling. Color Res. Appl. **29**(5), 381–389 (2004)
7. Ou, L.C., Luo, M.R.: A colour harmony model for two-colour combinations. Color Res. Appl. **31**(3), 191–204 (2006)
8. Geusebroek, J., Smeulders, A.: A six –stimulus theory for stochastic texture. IJCV **62**, 7–16 (2005)
9. Porter, M.F.: An algorithm for suffix stripping. Program **40**(3), 211–218 (2006)

10. The movie database. http://www.themviedb.org/
11. Subashin, K. Palanivel, S., Ramaligam, V.: Audio-Video based segmentation and classification using SVM. In: International Conference on Computing, Communication and Networking Technologies (2012)
12. Paris, G., Lambert, P., Beauchene, D., Deloule, F., Ionescu, B.: Animated Movie genre detection using symbolic fusion of text and image descriptors. In: International Workshop on Content-Based Multimedia Indexing, pp. 37–42 (2012)