Automatic Identification Genre of Audiovisual Documents

Manel Fourati, Anis Jedidi, Faiez Gargouri
Laboratory MIR@CL
University of Sfax.
Sfax, Tunis
Manel.Fourati@fsegs.rmu.tn, jedidianis@gmail.com, faiez.gargouri@isimsf.rmu.tn

Abstract—Identifying the Genre of an audiovisual document is among the major challenges for multimedia retrieval. Indeed, the lack of semantic metadata extraction makes these resources underused in the retrieval process. To overcome these difficulties, the extraction of semantic descriptions requires an analysis of the audiovisual document’s content. The automation of the process of describing audiovisual documents is essential because of the richness and the diversity of the available analytical criteria. In this paper, we present a method that allows the identification of a semantic and automatic description from the content such as genre. We chose to describe the cinematic audiovisual documents based on the documentation prepared in the pre-production phase of films, namely synopsis. The experimental result on Imdb (Internet Movie Database) and the Wikipedia encyclopedia indicate that our method of genre detection is better than the result of these corpuses.

Keywords—Identification, Audiovisual documents, Genre, Semantic.

I. INTRODUCTION

The quantity of audiovisual data available on the web and in audiovisual information databases is growing exponentially. Despite the use of adaptive techniques of information research in the exploration of audiovisual documents, the problem of the user’s dissatisfaction persists in different dimensions such as the exploitation and the research on audiovisual documents, especially the cinematic documents. To resolve this problem, we find it essential to extract some representative descriptions of content of cinematic documents such as genre descriptions.

Knowing that the genre represents a significant description for the films [1], we focus on the extraction of this description through a textual analysis from an unstructured textual document. The originality of the proposed method is that the extraction of the genre description is made in an automatic and semantic way.

This paper is organized as follows: The next section discusses the related works that deal with different techniques of genre and key word identifying of audiovisual documents. Section 3 is an overview of our approach of extracting genre. Section 4 presents the experiments of audiovisual genre detection. Finally, the last section concludes the paper and deals with future works.

II. RELATED WORKS

In order to identify the genre’s audiovisual documents, it is necessary to extract representative descriptions of the audiovisual document. The use of these descriptions is a necessary condition to reach the required information easily. In this section, we present an overview of the most relevant works proposed in literature related to the description and identification of audiovisual genre and the textual document genre.

A. Genre audiovisual identification

As part of the identification genre’s audiovisual documents, several approaches are based on the classification of the general genre, namely [2] and [3]. [2] used the weighted kernel logistic regression for video classification in the following genres: News, Sport and Movies. In order to extract the appropriate features, the authors extracted the key frames, used DCT (Discrete Cosine Transform) to obtain DCT coefficients and applied PCA (Principal Component Analysis) to select the pertinent DCT coefficients. However, [3] have used a set of computational features from visual characteristics (effects, motion and color) for video classification in the following genres: Sports, Music, news, cartoons and commercials.

Some works proposed methods to identify the genre’s content of the audiovisual document; significant ones are [4] and [5]. [4] suggested, for movie genre categorization, a method have used low level visual features from scene categorization. This method use movie trailers for video classification in the following genres: action, comedy, drama, and horror. Gregory pais mentioned in [5] the use of a symbolic fusion between the textual information (extracted from synopsis) and color and activity information (extracted from image) for identifying animated movie genre “Drama”. In our work, we focus on genre identification of films’ content. We used textual information (synopsis) to identify all genres of films.

B. Description of the textual document genre

Some works propose methods based on linguistic analysis and some others are based on statistic analysis. [6] proposed a method of the statistical discriminant analysis. The inputs of this method are the features extracted from the document such as a part of speech tagger and personal pronouns. The outputs are a set of discriminant functions
that distinguish between genres. To improve his results, Karlgren [7] uses other simple statistical features: sentence length, word length, syntactic complexity. In [8] in order to identify the genre, the authors use word frequency and frequency of punctuation marks. To predict the membership to the genre group, Stamatos applies the discriminant analysis used in the works of [6].

A more interesting method in the literature is the use of the statistic frequency of words in the text. [9] and [10] propose using a weighting method which is commonly used for information retrieval, the TF-IDF. [11] introduced the deviation formula of TF-IDF to Tf ratio and Idf ratio to obtain a set of training documents used for the statistic classifier Naive Bayesian. [12] describes an approach based on linguistic analysis. This method uses linguistic cues to identify generic and automatic genres. These cues are classified into four categories: structural cues (e.g. passive, nominalization, part-of-speech...), lexical cues (e.g. terms of address (Mr., Ms.), character-level cues (e.g. punctuation marks) and derivative cues (e.g. ratios and measures of variation). Brett uses the statistical technique LR (Logistic Regression) and neural network (single and multilayer perceptron) as computational methods for modeling a response using a binary logic function. In [13], the author identifies the genre by analyzing the linguistic content of words appearing in the transcripts of the audio tracks of video. The method defines stop words frequency as discriminant terms of genre by applying the metric TF-IDF (Term Frequency-Inverse Document Frequency) for the genre and not the document.

More recently, [14] has reported in his work the use of words frequency and genre frequency. He selected four genres (fantasy, Science Fiction, Philosophy and classical literature) and computed the frequency of these genres in the dictionary (GWFD) and the frequency of words in several books and then he built a set of frequency dictionary words (OWFD). The most frequent words of the dictionary are selected as pertinent. In order to identify genres, Hyoyoung selected four colors and drew ellipses on these colors to indicate the genre of each word. Though interesting, the methods presented in the literature are hampered by shortcomings. They provide linguistic and statistical analyses without considering the semantics contained in the document. Such semantic descriptions extracted from the content of the document become a necessary condition for linking the document content and description.

In this paper, we propose a method of genre detection of audiovisual documents using the pre-production documents. A more interesting method taking into account the statistic, the linguistic and the semantic analysis, which is described in the rest of this paper.

III. OVERVIEW OF THE APPROACH

To describe the genre of audiovisual document, different axes are used such as audio, video, soundtracks. The objective behind this description is a better identification and retrieval of the audiovisual document. We will concentrate on the pre-production documentation to extract automatic genre and key words of the audiovisual document. The synopsis is a pre-production document which contains a lot of important information. It represents a summary of the script and describes the outlines of the film’s story. Indeed, the more semantic description there is, the more the user's satisfaction is ensured. We will reserve this section to provide an overview of our proposed approach for the identification of genre and key words’ audiovisual documents.

To identify these descriptions, we are trying to conceive a system based on the extraction of semantics from the audiovisual document. The specificity of our annotation approach of audiovisual documents using the pre-production document is based on a combination of a statistical, linguistic and a semantic analysis. The entire process of our approach is automatic. Figure 1 describes this method, which consists of four phases that we can shortly explain as follows: 1) The Pre-treatment phase, 2) The Genres’ Extraction phase, 3) The Genres’ Identification phase, and 4) The Key words identification phase.

![Identification of Audiovisual Documents’ Description](image)

**A. The pre-treatment phase**

The first phase of our approach is a pre-treatment phase. The input of this phase is the synopsis and the output is the vector which contains the most frequent terms of the document (V). From the synopsis we extract the pertinent phrases synopsis using the tool KEA (Key phrase Extraction) [15]. When the key phrases are extracted, we eliminate all semantically insignificant terms (removing the stop words) and lemmatizing each term using the Stanford lemmatizer. Finally, we extract the vector which contains the most frequent terms of the document. This vector represents the starting point of the second phase of our process: the Genre s’ extraction phase.

**B. The Genres Extraction phase**

In our work we identify all cinematic genres (e.g. Action, Adventure, Drama, science-fiction, comedy...). Two different techniques are used to extract the genre description from the document: The statistical measure TF-IDF and the semantic similarity measures.

- The statistical measure TF-IDF used to evaluate the importance of a term contained in a document. We propose adapting the TF-IDF metric not only to the predefined genres but to their synonyms and
hypernyms which are extracted from the semantic lexical database WordNet.

- The semantic similarity measures used to compute the semantic distance between terms/genres. We exploit these measures to estimate the proximity of each term (ie the most frequent words) in the different concepts extracted (ie genre, synonym and hypernym). In literature, several metrics have been proposed. The typical measures that we decided to adopt are: the Jaccard [18], the Cosine [19], the Dice [20] and the Overlap [21] measures.

Consider a matrix, M, which contains the extracted genres with their synonyms and hypernyms. A directed acyclic graph, $G=(S, A)$ represents the result of this phase of our process. With $S=\{\text{vertices}\} = \{\text{TFIDF, Jaccard, Dice, Overlap, Cosine}\}$, $A=\{\text{edges}\} = \{\text{res tfidf, res jaccard, res dice, res overlap, res cosine}\}$, $res_i = \{\text{Ti, V, C}\}$, $\text{Ti}$=Term i, $V$= measured value and ‘C’= instance of genre. For each term $\text{Ti}$, we calculate the values for two different techniques previously cited (TF-IDF and The semantic similarity measures) to obtain the result res i. If the term $\text{Ti}$ has a non-zero value it will be stored in the graph $G$. Below, we present an example:

\[
\begin{align*}
\text{If } (\text{res tfidf}(\text{Ti}, M)! =0) & \text{ then } \\
G[1] & \leftarrow \text{add}(\text{Ti}, V, C) \\
\text{End if}
\end{align*}
\]

\[
\begin{align*}
\text{If } (\text{res cosine}(\text{Ti}, M)! =0) & \text{ then } \\
G[2] & \leftarrow \text{add}(\text{Ti}, V, C) \\
\text{End if}
\end{align*}
\]

C. The Genre Identification phase

The result from the genre extraction phase is the base core of the Genre Identification. The oriented acyclic graph $G$ represents all the genres automatically extracted from the document in the different techniques previously cited. The Following Figure shows an example of a modelling graph chained list.

![Fig. 1. example of a modelling graph](image)

In order to identify the key genres from the content of a document, the main idea is to find the related words for each genre. Each genre ‘C’ contains one or several terms ‘Ti’, and each term ‘Ti’ has at least one representative value. We propose measuring the pertinence of each genre. Two cases are then possible:

- $w_{Ti} = \max (v)$ If the term $\text{Ti}$ refers to the genre ‘C’ through more than one distance. Where: $w_{Ti}$ is the weight of the term $\text{Ti}$ in the genre ‘C’.

- $w_{Ti} = v$ If the term $\text{Ti}$ refers to the genre ‘C’ through only a single distance.

The pertinence of a genre ‘C’ is measured by the following formula:

\[
\text{pert}(C) = \frac{\sum_{i=1}^{n} w_{Ti}}{nb}
\]

where $nb$ represents the number of terms according to the genre ‘C’. Below, we present an example of pertinence computing.

We assume that $C1 = \text{Action}$ and as shown in Figure 1 the terms $T1$, $T2$ and $T5$ are related to $C1$ ie. $wT1= \max (v1, v2, v8, v10)$, $wT2=v4$ and $wT5= \max (v7, v11)$

\[
\text{pert}(c1)=( wT1+ wT2+ wT5)/3
\]

To sum up, our automatic system gives as a result some semantic genres extracted from the content. The genre having the highest pertinence represents the dominant genre in the description extracted and all the other concepts are considered as secondary genres.

D. The Key word identification phase

When the genres are extracted from a semantic and automatic process, we identify the key words related to the content of audiovisual document. Below, we present an algorithm which shows the steps of the extraction Key words description.

![Algorithm Key word identification](image)

IV. EXPERIMENTS:

We collected 150 English documents from the online database Imdb (Internet Movie Database) [22]. This base categorizes films in one to five genres from twenty-four genres. In our work, we use all genres of films (33 genres extracted from the Imdb and the Wikipedia encyclopedia). These web collections are used for testing and training the genre identification proposed in this paper.

In [23], we tested our experimental result in only 60 documents. We used criteria for measuring performance: exhaustivity and specificity. In this paper, we tested our experimental result in 150 synopses and we compared our approach with the result already mentioned in the Wikipedia
As shown in Table I, we compared our approach with the work presented in [5]. The values shown in the Table I represent the performance measures of its work using the synopsis of the film. The authors tested their work on 107 synopses and used only the genre “drama”. The results were: $P=43\%$, $R=81\%$ and $Fscore=46\%$. By comparing the performance measures of our test outcome with these values, we find that the experimental results are interesting, as 150 synopses are tested for all genres of movies.

V. CONCLUSION

This paper presents an automatic audiovisual documents genre and key words description. The objective of this approach is to overcome the semantic analysis gap in order to extract the audiovisual genre. In this context, we use the pre-production documents to combine the statistical analysis and the semantic analysis. Our statistical analysis is based on the TF-IDF. We propose adapting this metric to the semantic lexical database WordNet. In addition, we exploit four semantic similarity measures to estimate the proximity between terms and genres. In our future work, we plan to explore the structuring and the homogenization of the descriptions by XML descriptors (metadata) integrated in the MPEG7 standard.

VI. REFERENCES


