Automatic Audiovisual Documents Genre Description

Manel Fourati, Anis Jedidi and Faiez Gargouri
Laboratory MIR@CL, University of Sfax, Sfax, Tunisia

Keywords: Annotation, Audiovisual Documents, Genre, Semantic.

Abstract: Audiovisual documents are among the most proliferated resources. Faced with these huge quantities produced every day, the lack of significant descriptions without missing the important content arises. The extraction of these 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 extraction of a semantic and automatic description from the content such as genre. We chose to describe 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.

1 INTRODUCTION

By seeing the large amount of audiovisual documents produced each day, the exploitation and the research on audiovisual documents, especially cinematic documents, became a major issue that has grown significantly in recent years. 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 (Marc, N. et al., 2007), 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 description 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

2 RELATED WORKS

Describing is the process of extracting representative descriptions of the audiovisual document in order to obtain annotations. 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 of the document genre. Some works propose methods based on linguistic analysis and some others are based on statistic analysis.

(Karlgren, J. et al., 1994) 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 (Karlgren, J. et al., 1998) uses other simple statistical features: sentence length, word length, syntactic complexity.

In (E. Stamatatos et al., 2000) 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 (Karlgren, J. et al., 1994).

A more interesting method in the literature is the use of the statistic frequency of words in the text. (Brezealed, B. et al., 2006) and (Lin Wei-Hao et al., 2002) propose using a weighting method which is commonly used one for information retrieval, the TF*IDF. (Yong-Bae Lee et al., 2002) introduced the deviation formula of TF*IDF to Tf ratio and Idf ratio

Fourati M., Jedidi A. and Gargouri F..
Automatic Audiovisual Documents Genre Description.
DOI: 10.5220/0005170905380543
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (SSTM-2014), pages 538-543
ISBN: 978-989-758-048-2

to obtain a set of training documents used for the statistic classifier Naive Bayesian.

(Brett, K. et al., 1997) 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 (Stanislas, O. et al., 2010), 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, (Hyoyoung K. et al., 2013) 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.

3 OVERVIEW OF THE APPROACH

In this section, we give an overview of our proposed approach for the description of audiovisual documents. In general, the audiovisual document includes different axis to extract description, such as audio, video, soundtracks. As a consequence, when a user needs to extract generic descriptions like the genre, from documents, he needs to gather results in a short time without missing the important content. To resolve this problem, we will concentrate on the pre-production documentation to extract automatic genre 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.

To extract this description, 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. Using an interface, the user can access the database to select the synopsis of the film in order to extract the genres from the content.

Figure 1 describes this method, which consists of three phases that we can shortly explain as follows: i) The Pre-treatment phase, ii) The Description Extraction phase, and iii) Identification of the description phase.

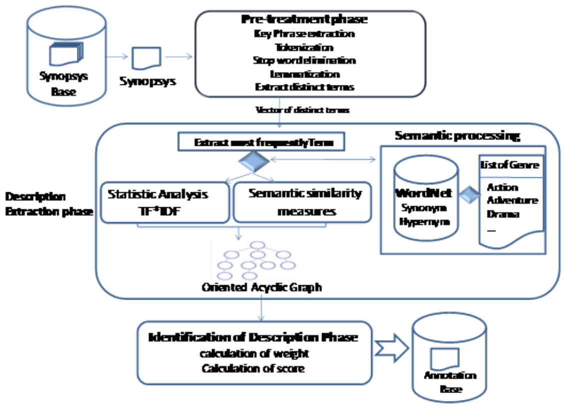

Figure 1: Description of Audiovisual Documents.

3.1 The Pre-Treatment Phase

As it is shown in Figure 1, the first phase of our process of description of audiovisual documents is a pre-treatment phase. This phase is based on five steps to obtain a vector of terms which contains significant terms: (1) Extraction of the pertinent phrases of the synopsis using the tool KEA (Key phrase Extraction) (Medelyan, O. et al., 2006). It extracts the most relevant sentences in the document. (2) Tokenization of these sentences (3) Eliminating all semantically insignificant terms' during which we remove the stop words. (4) Lemmatizing each term using the Stanford lemmatizer and (5) Extracting of the vector which contains the most frequent terms of the document. Knowing that, several works in literature have demonstrated the importance of the most frequently used words to analyze documents.

The resulting vector of this phase is the starting point of the second phase of our process: the description extraction phase.

3.2 The Description Extraction Phase

In this phase, we have used two different techniques when extracting annotation as it is shown in Figure 1. The first technique is the weighting method TF-IDF. This statistical measure is used to evaluate the importance of a term contained in a document, in relation to a collection or corpus (Ronan Cummins. 2013). In the work of (Stefano, C. et al., 2013), it identified the discriminant terms using the TF-IDF metric on predefined genres. In our work we identify all cinematic genres (e.g. Action, Adventure, Drama, science-fiction, comedy...). We propose adapting the TF-IDF metric not only for the predefined genres but for their synonyms and hypernyms which are extracted from the semantic lexical database WordNet.

In the second technique, we use the semantic similarity measures to generate descriptions. However, the semantic similarity measures represent a core base to compute the semantic distance between terms/genres. In literature, several metrics have been proposed. The typical measures that we decided to adopt are: the Jaccard (Jaccard p. 1901), the Cosine (C. Van Rijsbergen. 1979), the Dice (Nei, M., Li, WH. 1979) and the Overlap (LR Lawlor, 1980) measures. 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). If the calculated measure between the term and the genre admit a non-zero value, we find that there is a semantic relationship. Below, we present an algorithm which shows the steps of the two description extraction techniques previously mentioned.

Given that:

- M: is a matrix which contains the extracted genres with their synonyms and hyperonyms.
- G: is a directed acyclic graph, G=(S, A).
- S= {vertices} = {TFIDF, Jaccard, Dice, Overlap, Cosine}.
- A = {edges}
- = {res tfidf, res jaccard, res dice, res overlap, res
- cosine}.
- res i = {Ti, V, C}.
- Ti=Term i, V= measured value and 'C'= instance of genre.

Algorithm Description Extraction

```
Input: T= {T1, T2..., Tn}: a vector of
most frequent terms
Output: G: oriented acyclic graph
modeled in a chained list table
Begin
M [l, c]: a matrix having l rows and c
columns
n: size_T
G []←null
M← result of wordnet
For each i ∊ [1..n] do
If (res_TFIDF(Ti, M)! =0) then
G[1]← add(Ti, V, C) End if
If (res_cosine(Ti, M)! =0) then
G [2]←add(Ti, V, C) End if
If (res_Jaccard(Ti, M)! =0) then
G [3]←add(Ti, V, C) End if
If (res_Dice(Ti, M)! =0) then
G [4]←add(Ti, V, C)  End if
If (res_Overlap(Ti, M)! =0) then
G [5]←add(Ti, V, C) End if
End for
End
```

3.3 Identification of Description Phase

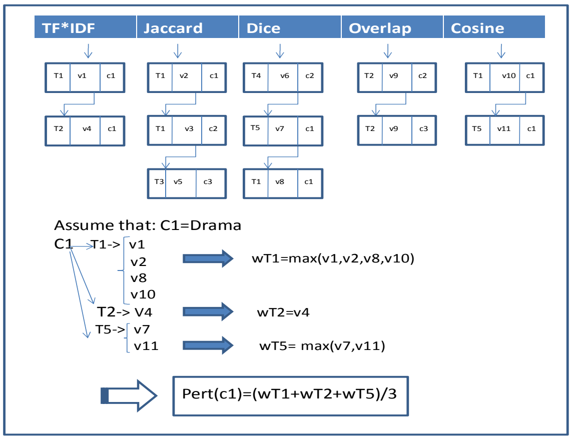

Figure 2: Example of Genre Identification.

The oriented acyclic graph G represents the result of the second phase of our process. Indeed, it represents all the genres automatically extracted from the document in the different techniques previously cited. In this section, we identify key genres that are closely related to the content of a document. The main idea is to find the related words for each genre. The Following Figure shows an example of a modelling graph chained list and calculates the relevance of a genre.

Each genre 'C' contains one or several terms 'Ti', and each term 'Ti' has at least one representative value. In order to identify these concepts, we propose measuring the pertinence of each genre. Two cases are then possible:

- If the term Ti refers to the genre 'C' through more than one distance then

$$wTi = \max(\mathrm{v}) \tag{1}$$

Where: w_{Ti} is the weight of the term Ti in the genre 'C'.

- If the term Ti refers the genre 'C' through only a single distance then

$$wTi = \mathrm{v} \tag{2}$$

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

$$pert(C) = \frac{\sum_{i=1}^{nb} \mathrm{wTi}}{nb} \tag{3}$$

where nb represents the number of terms according to the genre 'C'.

To sum up, our automatic system gives as 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.

4 EXPERIMENTS

Our experimental corpus is collected from the online database Imdb (Internet Movie Database). The total number of English documents collected is 60.These web collections are used to test the genre description proposed in this paper.

To evaluate our experiment results, we used criteria for measuring performance. For each document, we compared the genres extracted from the system with the result already mentioned in the Wikipedia encyclopedia and in the Imdb database. To estimate this pertinence, we used two parameters: exhaustivity and specificity (Stefano, C. et al., 2013). In this context, the likelihood ratio (Chernoff, H., 1954) is used to capture the degree of estimation of these two parameters. Table 1 displays the formula used to calculate the ratio Likelihood.

Table 1: The ratio Likelihood.

	Truth(+)	Truth(-)
Test(+)	a	b
Test(-)	c	d

Given that:

- a :the number of genres as test outcome is positive and true in test databases(Imdb, Wikipedia)
- b: the number of genres as test outcome is positive and false in test databases.
- c: the number of genres as test outcome is negative and true in test databases.
- d: the number of genres as test outcome is negative and false in test databases.

Where:

$$Se = \frac{a}{a+c} \text{ and } \mathrm{Sp} = \frac{d}{d+b} \tag{4}$$

$$\Rightarrow \mathrm{Lr(+)} = \frac{Se}{1-Sp} \text{ and } \quad \mathrm{Lr(-)} = \frac{1-Se}{Sp} \tag{5}$$

Table 2 is an example of calculation result of two parameters for documents (Lr(+) represents exhaustivity and LR(-) represents specificity):

Table 2: Example of five films' experiment result.

Film	Imdb		Wikipedia	
	Exhaustivity	Specificity	Exhaustivity	Specificity
A clock orange	16,5	0,35	14,28	0
Panoraman	7,5	0,43	25	0
looper	12,5	0,52	0	1,11
American beauty	25	0	25	0
Anna karenine	25	0	12.5	0.52

The likelihood ratio shows the degree of relevance of genres extracted from our system by mentioning two results: positive result (Lr+) and negative result (Lr-). The aim is to obtain a high degree of exhaustivity and a degree of specificity <1 (converging to 0).

Figure 3 show the experimental result of the specificity. The two curves displayed shows the LR-values by comparing our test outcome with to the outcome of the two bases before mentioned (wikipedia and imdb) for 60 films. We note that this values converging to 0.

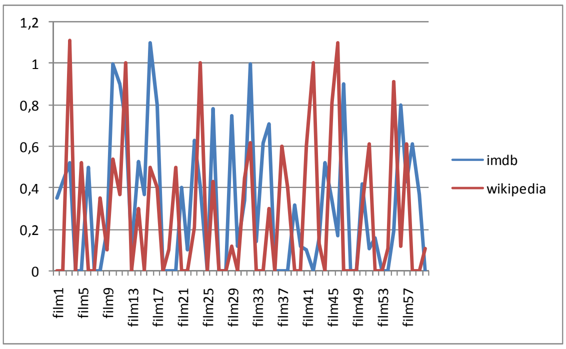

Figure 3: The experimental specificity result.

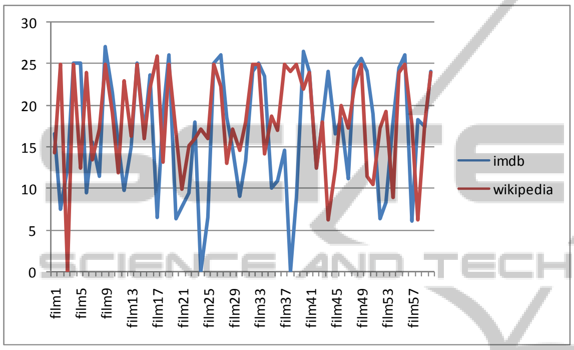

Figure 4: The experimental exhaustivity result.

Figure 4 show the experimental result of the exhaustivity. The two curves displayed shows the LR+ values by comparing our test outcome with to the outcome of the two bases before mentioned (wikipedia and imdb) for 60 films. We obtain a high degree of exhaustivity (the most values are between 10 and 25).

When we examine these measures calculated on the evaluation corpus, we note that the experimental results are interesting) by comparing the correct description rate of our test outcome (~83%) with the correct description rate to the work of (Stanislas, O. et al., 2010) (80%).

5 CONCLUSIONS

This paper presents an automatic audiovisual documents genre 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.

REFERENCES

Brezealed, B and Cook, D. 2006. Using closed captions and visual features to classify movies by genre. *In Proceedings of Multimedia Data Mining/Knowledge Discovery and Data Mining.* MDM '07, New York, NY, USA, Article No. 4.

Brett, K., Geoffrey, N., Hinrich, S. 1997. “Automatic Detection of Text Genre”. ACL’97, Madrid, pages 32-38.

Chernoff, H., 1954. On the distribution of the likelihood ratio. Ann. Math. Stat. 25:573–578. *Journal*,Volume 25, Number 3 (1954), 573-578.

C. Van Rijsbergen. 1979. Information Retrieval (Book 2nd ed). *Butterworth-Heinemann Newton, MA, USA*, 1979.

E.Stamatatos, N.Fakotakis, G. Kokkinakis, “Text Genre Detection Using Common Word Frequencies”, *In Proc. of the 18th Int. Conf. on Computational Linguistics* (COLING2000) Saarbrücken, Germany, pp. 808-814, 2000

Hyoyoung, K., Jin, Wan Park. , Genre Visualization Based on Words Used in Text. *International Conference, HCI International 2013*, Las Vegas, NV, USA, July, 2013, Proceedings, Part II .pp 551-554.

Internet Movie Database, Available at: http://www.imdb. com/ last visit:18/07/2014.

Jaccard p. 1901.Etude comparative de la distribution florale dans une portion des alpes et des jura. Bulletin de la société vaudoise des sciences naturelles, vol 37, pp 547-5790.

Karlgren, J and Douglass,C. 1994. Recognizing text genres with simple metrics using discriminant analysis. *In Proceedings of Coling 94*, Kyoto, Japan, pp 1071-1075.

Karlgren, J., Ivan, B., Johan, D., Anders, H., Niklas, W., Iterative Information Retrieval Using Fast Clustering and Usage-Specific Genres”, *8th DELOS workshop on User Interfaces in Digital Libraries,* Stockholm, Sweden, 21-23 October 1998. Pages 85-92.

Lin Wei-Hao and Alexander Hauptmann. 2002. News video classification using SVM-based multimodal classifiers and combination strategies. *In Proceedings of the tenth ACM international conference on Multimedia.* ACM, New York, NY, USA, 323-326.

LR Lawlor, Overlap, similarity, and competition coefficients. Journal, Vol. 61, No. 2, Apr., 1980.Pages 245-251.

Marc, N., Jocelyne N., Peter, R. King and Ludovic, G. 2007.Genre Driven Multimedia Document Production by means of Incremental Transformation. *In proceedings of the 2007 ACM symposium on*

Document engineering. ACM, New York, NY, USA, 111-120.

Medelyan, O., Witten I. H. (2006) "Thesaurus Based Automatic Keyphrase Indexing." *In Proc. of the Joint Conference on Digital Libraries* 2006, Chapel Hill, NC, USA, pp. 296-297.

Nei, M., Li, WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA 76*: 5269—5273.

Ronan Cummins. 2013. A Standard Document Score for Information Retrieval. *In Proceedings of the 2013 Conference on the Theory of Information Retrieval*, Oren Kurland, Donald Metzler, Christina Lioma, Birger Larsen, and Peter Ingwersen (Eds.). ACM, New York, NY, USA, Pages 24, 4 pages.

Stanislas, O., Mickael, R. and Georges, L. Transcription-based video genre classification. IEEE. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 14-19 March 2010. Dallas, TX. Pages 5114 – 5117.

Stefano, C., Alessandro, B., Marco, B., Emanuele, DV., Piero F., Silvia, Q. 2013. An Introduction to Information Retrieval. *Book* Web Information Retrieval, 2013. pp 3-11.

Yong-Bae Lee and Sung Hyon Myaeng (2002). Text Genre Classification with Genre-Revealing and Subject-Revealing Features*. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 2002. ACM, New York, NY, USA. Pages 145-150.