

Discovering Indexing Rules for Video-News

Marcin DETYNIĘCKI
CNRS, Pole IA, LIP6
Université P. et M. Curie
8, rue du Capitaine Scott
75015 Paris - France
+ 33 1 44 27 88 03
Marcin.Detyniecki@lip6.fr

ABSTRACT

In this paper, we present how we discover three indexing rules for video-news. These rules are intended for finding the structure (macro-segmenting) the video-news. We use a fuzzy decision tree to learn this rules directly from simple color proportions extracted from key-frames taken from the video. Experimental results and comparisons are presented.

Keywords

Data Mining, Video, Fuzzy Decision Trees.

1. INTRODUCTION

On the one hand, the growth of video data has caused a corresponding need to analyze and exploit it. Perfect examples of this increase are the availability of video news on the web, or the increase of home video storage. Traditionally, it appears that the user tends to interact with this kind of objects in order to obtain what he wants. But in order to provide what the user wants indexing is needed.

However, indexing is generally done manually. The growth of video data, the personal storage aspect and the demand of new applications for finer grain access, highlight the fact that automation of the indexing process is needed.

On the other hand, in the recent years, knowledge discovery from data and data mining introduces new methodologies to extract and discover automatically knowledge from data repositories.

Thus, it appears natural and promising to link data mining with multimedia data that leads to multimedia mining. In fact, the extracted knowledge can be used directly as index or even better to improve the indexing process.

Dealing with multi-media introduces a new problem related to the polymorphism of the data [2]. In fact the information that can be extracted from a video are texts, images, sounds, temporal data, metadata, etc. A solution lies in providing a flexible and automated data-mining tool, which will induce knowledge from all kinds of data. A particular instance of such tools is the fuzzy decision tree based algorithm [1].

The structure of this paper is the following. In Section 2, we shortly introduce and provide some references of our data-mining tool: the fuzzy decision trees. Section 3 is the core of the paper. In this section we illustrate the knowledge discovery process. We extract different type of rule rules based on the color feature of per-shot key-frames. We finish this section with a short discussion.

2. FUZZY DECISION TREES

Knowledge Discovery from Data (KDD) was introduced by Fayyad, Piatetsky-Shapiro and Smyth at the beginning of the nineties [3]. Due to the singular difficulty of the multimedia case (for instance due to the polymorphism of data) a new topic of research was recently proposed: Multimedia Data Mining (MDM) [4].

Therefore in a multimedia framework versatile data-mining tools are necessary. One particular case of these tools is the fuzzy decision tree learning algorithm, which provides rules that summarize and explain the data. The used software (Salammô) is able to handle numerical input, and constructs fuzzy decision trees [1].

The advantage of using fuzzy decision trees resides in the fact that they represent a natural way understandable knowledge. A fuzzy decision tree is equivalent to a set of fuzzy "if...then" rules, where the premises, conclusions and variables can easily be interpreted.

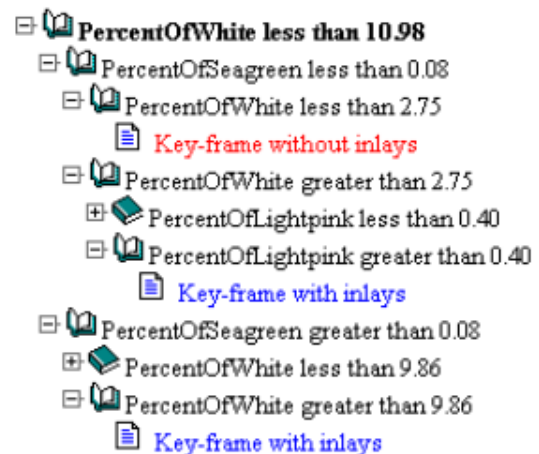


Figure 1. Example of rule extracted by Salammô.

3. DISCOVERING INDEXING RULES

In order to show the interest of using fuzzy trees (and more generally any data mining tool), we reduced our research to only one well-known feature: the color.

We start from a set of per shot key-frames extracted from the video flow [5]. Here, a key-frame is an image in GIF format each with 256 specific colors. In a second step, the set of colors of each key-frame is vectorized and "projected" into a given reference-

palette (for instance a palette of 64 or 256 colors obtained by discretizing equally the RGB space). We obtain like this a common basis to compare all key-frames. A frequency histogram is then computed for each key-frame, which is then represented as a vector in the reference space (defined by the colors of the reference palette).

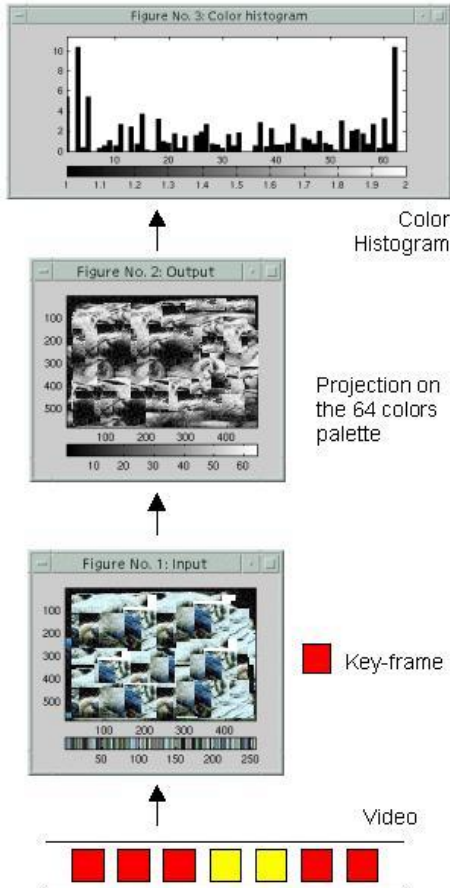


Figure 2. Color histogram extraction. (not actual key-frames)

3.1 Experimentations

As we just explained, each key-frame enables us to construct a vector, which can be considered as a training example. Thus, from an original set of key-frames, a training set can be composed to feed the Salammbo software and give rise to a fuzzy decision tree.

In order to show the versatility of the approach, three different mining problems have been considered (all related to the extraction of the structure based on the colors).

3.1.1 Discovering the presence of inlays

Inlays that appear on the TV screen are very often hints for the structure of the video news. They often appear when a new person is presented or when a report related to a subject ends (for instance presenting the name of the author of the report). They usually consist in a simple form (usually a rectangle) that frames some text.

We have conducted several mining experiments in order to determine if colors were discriminant for the detection of inlays. Our training set was composed with 176 vectorized key-frames, and each vector was assigned to a class (key-frame with or without inlays).

A first experiment was conducted with the whole training set and the reference palette composed by 64 colors. The resulting fuzzy decision tree is not very deep and has a root node on which the presence of the white color is asked (white is the major background color of the inlays in our training set). This result points out that only a few number of colors is needed to discriminate inlay key-frames. And more generally, we confirm the observation that the use of colors is a suitable feature for discriminating inlays. If we look closer at the constructed tree we notice that this intuitive rule is completed by the fact that only an intermediate proportion of this white color is indicative of the presence of inlays. Too much white or just very little are usually not frames with inlays.

A second experiment was conducted to position the Salammbo algorithm with regard to other learning algorithms (see table 1). The tests have been done thanks to the free software Weka [6]. We insist here on the fact that in this kind of applications, the recall and precision values are as important as the accuracy of the model. In fact, it is important to perfectly recognize at least one kind of key-frames.

We can observe that in addition to the understandability of the fuzzy decision tree model, Salammbo is not only among the ones with the higher accuracy, but also, presents high recall and precision rates for the inlay recognition. Moreover, it appears that the construction of fuzzy decision trees by Salammbo is also among the low time-consuming methods (an important property for multimedia applications).

3.1.2 Discovering errors in the shot detection

Shot detection algorithms are usually very sensitive. The idea is to be sure not to miss any shot change. The drawback of this approach is the over-segmentation of the video.

Thus, we have conducted an experiment to discover if based on the colors we can tell that two successive key-frames are from the same shot. Here, the training set was constructed as follows. A first group of training examples was composed by two successive key-frames from the same shot. A second group was composed by two successive key-frames from different shots. Each key-frame was vectorized in a 64 colors palette, the two vectors were merged to obtain a single training example with 128 features. In other words we just feed the fuzzy tree with a vector having the proportions of all the colors and no other underlying knowledge. For instance we did not provide any knowledge about the relationship between color i of the first palette and the same color i of the second.

At this point, we expect to find rules translating the color similarity. But, the built fuzzy decision tree highlights that the most discriminant approach is to detect the increase of the proportion of a particular color: the dominant color of inlays.

This rule suggests in the one hand that an often cause of error is the appearance of inlays. In the other hand that after a classical shot detection (based on similarities), the accuracy of the shot detection can be enhanced using the color dissimilarity between

Table 1. Results for inlays recognition (64-colors palette)

Algorithm	Accur. (%)	With inlays		Without inlays		Bld. Time (s)
		Recall	Precision	Recall	Precision	
Salambô (FDT)	81.3	0.88	0.78	0.75	0.86	1
Naive Bayes	54.6	0.39	0.57	0.71	0.53	0.4
Voted Perceptron	71.6	0.71	0.72	0.73	0.71	0.6
Weka J48 (C4.5)	78.4	0.75	0.81	0.82	0.77	1
Decision table	82.9	0.87	0.81	0.8	0.85	3.3
Adaboost (J48)	85.8	0.87	0.85	0.85	0.86	5.6
Neural Network	80.1	0.84	0.78	0.76	0.83	322

successive key-frames. In this experiment, the mean accuracy with cross-validation of fuzzy decision trees is 78.26%.

3.1.3 Discovering the host

A crucial hint about the structure of the news is to detect the appearance of the host. As in previous experiments, a training set was constructed. A first group of training examples was composed by key-frames where the host appears. A second group was composed by key-frames without host. Each key-frame was vectorized in a 256 colors palette.

The most effective fuzzy rule to recognize if we are in presence of the host (or not) is based on one color of the host's background (presently blue). Surprisingly, this rule points out that the best way to know if we are in presence of the host is to look at the background (i.e. if the scene takes place in the channel studio).

The accuracy of the fuzzy decision tree here is 88%.

3.1.4 Discussion

For these three different problems, the extracted knowledge is in the form of three unexpected seminal rules. Note that we did not introduce any a priori knowledge.

- In order to recognize the presence of inlays, the system suggests detecting a quite large proportion of a single color (no to large and not to small), putting forward that all inlays are large forms with uniform colors.
- In order to ameliorate the shot detection (which is naturally based on similarity) it is recommended to look at the differences between key-frames. More precisely, the rule suggests that an inlay has been detected, implying at the same time that this is actually the cause of the errors.
- In order to detect the presence of a host, focus on the background, which corresponds to discern if the scene was taken in the studios. So, this rule suggests not looking at the host to detect the presence of a host.

4. CONCLUSION

In this paper, we presented an example of multimedia knowledge discovery applied to the video-news structuring. We used fuzzy decision trees, because of the simplicity and the understandability of the extracted rules.

We focused on the mining of the color feature of per shot key-frames. The extracted rules provide hints for better indexing

in a macro-segmentation perspective. In fact, these rules deal with the appearance of important information on the screen (inlays), the presence of the host, or correcting possible mistakes in the shot detection.

This is a first step in the multimedia mining of a video, which shows its potentiality. The next step is to continue the mining of visual contents as for instance the texture and also the mining of structural content. Future work will consider the other medias (sound, text, etc.) in order to exploit the interaction between medias.

5. ACKNOWLEDGMENTS

This work was partially founded by the European project KLIMIT: Knowledge InterMediation Technology (ITEA 00008).

6. REFERENCES

- [1] C. Marsala and B. Bouchon-Meunier, An Adaptable System to Construct Fuzzy Decision Trees in Proceedings of NAFIPS'99 (New-York, June 1999), pp. 223-227.
- [2] M. Detyniecki and C. Marsala, Fuzzy inductive learning for multimedia mining in Proceedings of EUSFLAT'01 (Leicester, UK, Sept. 2001), pp. 390-393.
- [3] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, From Data Mining to Knowledge Discovery in Databases in AI Magazine, 17:3, pp. 37-54, 1996.
- [4] First Workshop of Multimedia Data Mining. http://www.cs.ualberta.ca/~zaiane/mdm_kdd2000/
- [5] R. Ruiloba and P. Joly, Framework for evaluation of video-to-shots segmentation algorithms in Video Data special issue of Networking and Information Systems, Vol. 3, pp. 46-57, 2000.
- [6] Weka 3 - Machine Learning Software in Java. <http://www.cs.waikato.ac.nz/~ml/weka>