Fuzzy Data Mining for Video

Christophe Marsala
Pôle IA, LIP6
Université P. et M. Curie - Paris 6
8 rue du Capitaine Scott,
75015 Paris, France
Christophe.Marsala@lip6.fr

Marcin Detyniecki
CNRS, Pôle IA, LIP6
Université P. et M. Curie - Paris 6
8 rue du Capitaine Scott,
75015 Paris, France
Marcin.Detyniecki@lip6.fr

Abstract

In this paper, we present a fuzzy data mining approach to mine video. Two experiments are presented to detect inlays in video by means of a fuzzy decision tree.

Keywords: Fuzzy data mining, Multimedia systems, fuzzy decision tree, video mining.

1 Introduction

The management of large amount of data is a crucial task. Data Mining (DM), and Knowledge Discovery (KDD) are concerned with the extraction of information, or knowledge, from data provided by databases. A current issue in KDD is *multimedia mining* [17]. Its aim is to handle data with a complex structure such as, for instance, images, text, spatial data, multimedia data, or video. In this paper, we focus on the particular multimedia case of *video mining*, concerned with extracting knowledge from videos.

Video mining lies at the crossroads of video indexing and KDD. It brings tools from these domains to extract some knowledge from video, helping to understand its structure, and offering tools for automatic indexing. Video indexing is concerned with improving the access to particular parts of a video by generating and using some meta-data [6]. Video mining aims at introducing training capabilities from KDD into such a process.

Soft computing techniques have already been introduced in KDD, for instance the so-called *Fuzzy Data Mining* (FDM) [1, 11]. FDM brings robustness in KDD, and enhances the understandability of the extracted knowledge. It is appropriated to handle numerical data from the real world (*eg.* time, length,...) and induced symbolic or fuzzy representation for such data that enables the user to better understand the results of data mining. Thus, FDM is very well adapted to video [2, 3].

In this paper, we present a FDM approach to mine video and extract knowledge from news broadcasts. In Section 2, we present the features used to describe video frames and we the recall basis of the fuzzy data mining algorithm used in the experiments we propose. In Section 3, two experiments conducted on a video are described.

2 Learning from a Video

When handling a video, the main part lies in defining features to describe it.

2.1 Features of Video Frames

A lot of information is embedded in the format of video but it is very difficult to use it directly. For instance, the MPEG standards (MPEG-1, MPEG-2, and MPEG-4) are commonly used to code video information. These standards derive benefit from the redundancy of the video information to compress and code it more efficiently. Roughly speaking, a video is a stream of consecutive *frames*, combined with audio information (spoken text, music, sounds). The frame stream contains a large set of frames, for instance, television programs are coded by around 25 to 30 of such frames per second, which represents around one million frames for one hour of TV program. After decoding, each frame can be viewed as an image in JPEG, or GIF format, using 256 colors of size around 320x240 pixels.

Compression of a video can be done by selecting a subset of frames taking into account the fact that consecutive frames do not much differ. Consecutive frames are compared and similar ones are grouped, and a *key frame* is chosen to represent this set of similar frames (a *shot*).

Shot detection is a crucial task when working with video. Several shot detection algorithms already exist and perform well for various video programs (commercials, video clips, news broadcasts,...) [13]. A lot of them used the MPEG coding information to detect differences between successive key frames. A difficulty when detecting shots lies in the sensitivity of the detection algorithm that can detect too more shots than required. Another difficulty lies in handling "artistic" video programs where special movements of the camera have to be considered (for instance *fade in*, or *fade out*).

Classical pattern recognition techniques can be applied on each key frame from a video in order to extract features. For instance, the color histogram is a currently used feature.

2.2 Fuzzy Data Mining

Fuzzy Data Mining is an active domain of research. The power of the Fuzzy Set Theory when handling real and numerical data is appreciated and brings robustness and understandability in the KDD process. A well-known FDM algorithm is based on the construction of Fuzzy Decision Trees (FDT). FDT is the extension of classical decision trees to handle numerical, and fuzzy data [1].

Most algorithms to construct FDT proceed in the same way: the so-called *Top Down Induction of Decision Tree* (TDIDT) method. They build a tree from the root to the leaves, by successive partitioning the training set into subsets. Each partition is done by means of a test on an attribute and leads to the definition of a node of the tree. An attribute is selected thanks to a *measure of discrimination* H. Such a measure enables us to order the attributes according to an increasing accuracy when splitting the training set. The discriminating power of each attribute is valued with regard to the classes. The attribute with the highest discriminating power is selected to construct a node in the decision tree and to split the training set into subsets.

On each subset, the same process is applied until a stopping criterion is fulfilled that enables us to construct a leaf of the tree. For instance, if all the examples in the subset have the same class, then the stopping criterion is fulfilled. Another stopping criterion should be based on the measure of discrimination: if the value of the measure for the given subset is below a given threshold, then the stopping criterion is fulfilled.

Methods of constructing decision trees, whether crisp or fuzzy, differ mainly in their choice of H [10, 12]. Two well-known measures of discrimination are the entropy of fuzzy events (a fuzzification of Shannon entropy [1]) and the measure of ambiguity [16].

In this paper, we consider the Salammbô software [8] that implements a FDT construction algorithm. We restrict ourselves to the two well-known methods to construct FDT implemented in the Salammbô software: the fuzzy entropy based method, and Yuan & Shaw's method [16]. More details and a comparison of these two methods can be found in [9].

Moreover, an automatic method to construct a fuzzy partitions on a set of numerical values is implemented in the Salammbô software. When no expert knowledge is available to give fuzzy labels for numerical attributes, this automatic method enables us to find the best fuzzy partition associated with the training set of values (for more details, see [7]).

3 Experiments

In this paper, we present an application of a FDM algorithm, implemented in the Salammbô Software [8], to extract knowledge from a video. A software due to the multimedia group of LIP6 [4, 14] was used to cut a thirty-minute tape (an evening broadcast news by a French TV channel) into shots. Thus, a set of 471 key frames, each one representing a shot, was obtained.

Each key frame enables us to construct a case for training and testing the FDM algorithm. The set of features associated with each key frames is:

- its RGB color histogram on a reference palette of 64 colors (see [3]),

- its *video features* that take into account the properties of the shot the key frame represents: the running time of the shot, the kind of transition between the previous shot and this one (CUT or FADING) (this transition is detected by the shot detection software), and the running time of this transition.

We focus on experiments to detect key frames in which an inlay appears. From the whole set of key frames, a training set has been constructed by selecting 176 key frames: 88 contains an inlay, and 88 are normal key frames (without inlay). Our goal here is to obtain a training set with an equal proportion of key frames with and without inlay. A cross validation was conducted by splitting the training set into 4 subsets, each one with 22 examples of each class. A training was done using simultaneously 3 of these 4 sets, the fourth one used as a test set, and so on. At the end, the mean value of the obtained results is given.

3.1 Lonely key frame

In a previous work [3], we focused on experiments that only take into account the color feature. Now, we have conducted new experiments taking into account the whole set of features (color and video features) for each key frame. A comparison between these 2 approaches is given in Table 1.

In this table, we present the accuracy of the FDT, and their recall and precision rates for each class (with and without inlay). In the first line, *Salammbô (IPMU'02)*, the results obtained in [3] are recalled. In Line 2 (resp. Line 3), the results obtained with FDT constructed with the fuzzy entropy (resp. Yuan & Shaw's measure) are given. In Line 4 (resp. Line 5), the results obtained with FDT constructed with the fuzzy entropy (resp. Yuan & Shaw's measure) and a threshold as stopping criterion used to prune the FDT during the construction are given [1].

We can observed that taking into account video features does not improve the accuracy of the detection when considering the construction of FDT by means of the fuzzy entropy. In this case, Yuan & Shaw's measure gives more efficient trees.

The pruning of the FDT by means of a threshold greatly enhanced the accuracy (the induced FDT are more general in this case than without a stopping threshold). Moreover, in this case, the FDT constructed by these two methods are the same: the attributes selected near the root of the tree are the same with the two measures [9].

3.2 Dual key frames

We have conducted other experiments in order to study the benefit to consider the position of key frames in the video stream. Our hypothesis here is that successives key frames are highly linked.

A training case was constituted here by a pair of two consecutive key frames (the so-called *dual key frames*). For one key frame (simply called the *current key frame*), we will try to predict if it contains or not an inlay. The other key frame (called the *previous key frame*) will be used to help the prediction of the current one. The features of a training case are composed of the video features of the previous key frame, the video features of the current key frame, the difference of running times between the previous key frame and its predecessor, the difference of running times between the current key frame and the previous one, and for each color of the reference palette of colors, the difference between the proportion of this color in the current key frame and in the previous one. This represents a vector of 64 numerical values.

In Table 2, the accuracy rates for the Salammbô algorithm and for other learning algorithms[1] for the detection of inlays are given. We also present their recall and precision rates for each class (with and without inlay), and the building time (in seconds) of the model when considering the whole training set.

We can observe that all the algorithms do not de-

[1]These tests have been done by means of the free software Weka from the University of Waikato (NZ) (http://www.cs.waikato.ac.nz/ ml/weka/).

Table 1: Detection of inlays.

Method	Accuracy	With inlay		Without inlay	
		Recall	Precision	Recall	Precision
Salammbô (IPMU'02)	*81.25%*	*0.875*	*0.778*	*0.75*	*0.857*
Salammbô (Fuzzy Entropy)	76.14%	0.75	0.767	0.773	0.756
Salammbô (Yuan & Shaw)	82.39%	0.886	0.788	0.761	0.87
Salammbô (Fuzzy Entropy + thld)	85.8%	0.898	0.832	0.818	0.889
Salammbô (Yuan & Shaw + thld)	85.8%	0.898	0.832	0.818	0.889

Table 2: Detection of inlays by using dual key frames.

Method	Accuracy	With inlay		Without inlay		Build.
		Recall	Precision	Recall	Precision	time
Salammbô (Fuzzy Ent.)	69.89%	0.67	0.711	0.727	0.688	0.9
Salammbô (Y. & S.)	70.45%	0.682	0.714	0.727	0.696	1
Salammbô (Fuzzy Ent. + thld)	71.02%	0.693	0.718	0.727	0.703	0.4
Salammbô (Y. & S. + thld)	73.86%	0.693	0.763	0.784	0.719	0.4
Bagging (with J48)	75.0%	0.739	0.756	0.761	0.744	6.5
Decision Table	72.73%	0.773	0.708	0.682	0.75	15
AdaBoost (with J48)	71.02%	0.682	0.723	0.739	0.699	9.5
Decision trees (Weka J48)	70.45%	0.773	0.68	0.636	0.737	1.4
Voted Perceptron	61.36%	0.614	0.614	0.614	0.614	0.4
Naive Bayes	59.66%	0.636	0.589	0.557	0.605	0.4
Neural Networks (Weka)	57.39%	0.625	0.567	0.523	0.582	386

rive benefit from the new features of the training cases. The accuracy is lower than in the previous experiment. In general, the FDT are more efficient when detecting the presence of an inlay (the precision is high). However, they often misclassify key frames by considering they do not contain an inlay (the recall value is low). This means that the information provided by the previous key frame is not necessarily convenient to predict the presence of an inlay. The features of the previous one act as noise when predicting the class of the current key frame.

The root attributes for the constructed FDT are either the running time of the previous key frame, or the difference of running time between the current key frame and the previous one.

Here are some examples of fuzzy rules deduced from the FDT constructed by the Salammbô software, with Yuan & Shaw's measure, for the whole training set:

- If the running time of the previous key frame is short (fuzzy set: *less than 2 sec.*), and the difference of proportions for the color dark-green is low (*less than -75%*), then the current key frame does not contain an inlay.
- If the running time of the previous key frame is high (*greater than 2 sec.*), and the difference of proportions for the color brown-red is also high (*greater than 8%*), then the current key frame contains an inlay.

Moreover, we can observe the efficiency of the FDT compared to other algorithms. A meta-learner algorithm (bagging) is the only more accurate algorithm for this experiment. Finally, the building time of the FDT, which is a crucial parameter when handling video, can be considered as good.

4 Conclusion

In this paper, we present fuzzy data mining for video. Two experiments have been conducted to detect inlays in video key frames. On the one hand, we observed the importance of video features to improve the detection of inlays. On the other hand, the merging of features from a pair of consecutive key frames does not enhance the detection.

In future work, the classification of the previous key frame will be added to help the prediction of the current key frame. Moreover, the use of dual key frames will be studied in other kinds of applications. Other soft computing techniques will also be studied in order to compare key frames (see for instance [15]) or to construct fuzzy association rules (see for instance [5]) from the set of key frames .

Acknowledgements

We express our thanks to R. Ruiloba and G. Durand for providing the indexed record of video news.

References

[1] B. Bouchon-Meunier and C. Marsala. Learning fuzzy decision rules. In D. Dubois J. Bezdek and H. Prade, editors, *Fuzzy Sets in Approximate Reasoning and Information Systems*, volume 3 of *Handbook of Fuzzy Sets*, chapter 4. Kluwer Academic Publisher, 1999.

[2] M. Detyniecki and C. Marsala. Fuzzy inductive learning for multimedia mining. In *Proc. of the EUSFLAT'01 conference*, pages 390–393, Leicester (UK), September 2001.

[3] M. Detyniecki and C. Marsala. Fuzzy multimedia mining applied to video news. In *Proceedings of the IPMU'02 Conference*, pages 1001–1008, Annecy, France, July 2002.

[4] G. Durand and P. Faudemay. Cross-indexing and access to mixed-media contents. In *Proc. CBMI'01 - International Workshop on Content-Based Multimedia Indexing*, Brescia, Italy, Sept 2001.

[5] J. Kacprzyk and S. Zadrozny. Linguistic summarization of data sets using association rules. In *Proc. of the IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE*, pages 702–707, St Louis, USA, May 2003.

[6] S. Marchand-Maillet. Content-based video retrieval: An overview. Technical report, Computer Vision Group, University of Geneva, 2000.

[7] C. Marsala. Fuzzy partitioning methods. In W. Pedrycz, editor, *Granular Computing: an Emerging Paradigm*, pages 163–186. Springer-Verlag, 2001.

[8] C. Marsala and B. Bouchon-Meunier. An adaptable system to construct fuzzy decision trees. In *Proc. of the NAFIPS'99 (North American Fuzzy Information Processing Society)*, pages 223–227, New York, USA, 1999.

[9] C. Marsala and B. Bouchon-Meunier. Choice of a method for the construction of fuzzy decision trees. In *Proc. of the IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE*, pages 584–589, St Louis, USA, May 2003.

[10] C. Marsala, B. Bouchon-Meunier, and A. Ramer. Hierarchical model for discrimination measures. In *Proc. of the IFSA'99 World Congress*, pages 339–343, Taiwan, 1999.

[11] S. Mitra, S.K. Pal, and P. Mitra. Data mining in soft computing framework: A survey. *IEEE Transactions on Neural Networks*, 13(1):3–14, 2002.

[12] A. Ramer, B. Bouchon-Meunier, and C. Marsala. Analytical structure of hierarchical discrimination. In *Proc. of the 8th IEEE Int. Conf. on Fuzzy Systems, FUZZ-IEEE*, pages 1050–1053, Seoul, Korea, 1999.

[13] M. Roach, J. Mason, L.-Q. Xu, and F. Stentiford. Recent trends in video analysis: a taxonomy of video classification problems. In *Proc. of the 6th IASTED Int. Conf. on Internet and Multimedia Systems and Applications*, Hawai (USA), August 2002.

[14] R. Ruiloba and P. Joly. Framework for evaluation of video-to-shots segmentation algorithms. *Video Data special issue of Networking and Information Systems*, 3:46–57, 2000.

[15] D. Van de Weken, M. Nachtegael, E. Kerre, and G. Chen. Image quality evaluation.

In E. Damiani, R.J. Howlett, L.C. Jain, and N. Ichalkaranje, editors, *Knowledge-Based Intelligent Information Engineering Systems and Allied Technologies*, pages 326–330, 2002. Proc. of KES'2002 conference.

[16] Y. Yuan and M.J. Shaw. Induction of fuzzy decision trees. *Fuzzy Sets and systems*, 69:125–139, 1995.

[17] O. R. Zaiane, J. Han, Z.-N. Li, S. H. Chee, and J. Y. Chiang. MultiMediaMiner: A system prototype for MultiMedia data mining. *Proc. of the ACM Sigmod, Int. Conf. on Management of Data*, 27(2):581–583, 1998.