

# Fuzzy Inductive Learning For Multimedia Mining

Marcin Detyniecki

Berkeley initiative in Soft Computing (BISC)  
Computer Science Division- Department of EECS  
University of California, Berkeley, CA 94720  
marcin@eecs.berkeley.edu

Christophe Marsala

LIP6, University Paris 6  
8 rue du Capitaine Scott,  
75015 Paris, France  
Christophe.Marsala@lip6.fr

## Abstract

The growing of multimedia data has caused a corresponding growth in the need to analyze and to exploit it. In this paper, we propose to extract knowledge from the whole kind of multimedia data and to bring up a collaborative fuzzy data mining process where each kind of data helps to extract a global knowledge.

**Keywords:** Fuzzy data mining, Multimedia systems.

## 1 Introduction

In the recent years, fuzzy data mining introduces new methodologies to extract and discover fuzzy knowledge from either classical or fuzzy databases. It leads to the improvement of the knowledge discovery process that enables us to offer more comprehensive discovered knowledge, and to enhance its capabilities to handle real world data. Nowadays, a growing kind of databases handles a very rich form of data: textual, image, video, metadata,... A perfect example of multimedia data is a web page or a video, because they contain several kind of data: text, links, metadata, image...

Multimedia databases represent an important source of potential knowledge which is highly interesting for us to use. However, extracting knowledge from these databases is very difficult since on the one hand multimedia data are much richer than simple text and on the other hand handling such amount of information is already a

difficult task for a system. A solution lies in automated data mining which is a very interesting process to induce knowledge from databases [4].

Multimedia mining (MM) is a relatively new topic that stands for the mining of various types of data [7, 10, 14]. The aim here is to find knowledge by analyzing automatically all the existing kinds of data. MM will offer promising opportunities to handle various kind of data. MM is different from information retrieval since its aim is to find new knowledge rather than a specific information [8].

In this paper, we first recall in Section 2 the definition of fuzzy inductive learning in the knowledge discovery scheme. After that, in Section 3, we recall some basis on multimedia systems, and we propose some hints on how to obtain knowledge from raw fuzzy multimedia data in Section 4.

## 2 Fuzzy inductive learning and data mining

Data mining (DM) was first introduced at the beginning of the 1990s. The definition of knowledge discovery from data (KDD) has been introduced [4] as: "..., KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. DM is the application of specific algorithms for extracting patterns from data". Thus, data mining is the learning step in the KDD process where induction from data is conducted.

Inductive machine learning is a well-known topic with a large set of methods when the data to handle are symbolic ones. However, difficulties appear when considering data described by means

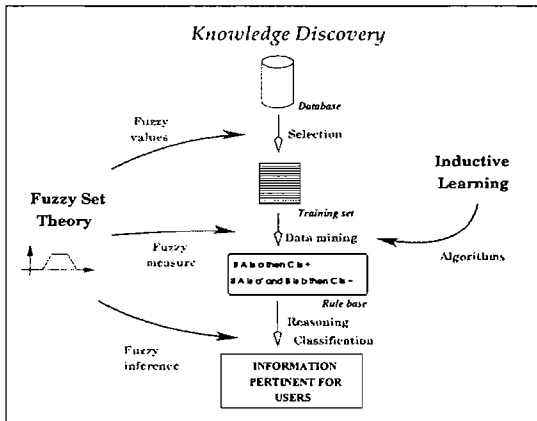


Figure 1: Fuzzy Knowledge Discovery

of fuzzy values. The introduction of fuzzy set theory in a learning method enables us to handle such values. For instance, this leads us to the construction of a set of fuzzy rules by means of the construction of *fuzzy decision trees* [1]. Each path in the tree is associated with a rule, where premises are composed by the tests of the encountered nodes, and the conclusion of the rule is composed by the class associated with the leaf of the path. Edges can be labeled by so-called fuzzy values in order to constitute a fuzzy rule base.

### 3 Multimedia systems

Multimedia systems are computer-delivered electronic systems that allow the user to control, combine, and/or manipulate different types of media (text, sound, video, graphics, animation,...).

One of the major difficulties that these systems face is to deal with different forms of data (and formats), as for instance text, hyperlinks, audio, images, video. Probably the most complete and at the same time complex media is the video. We note here that information is not directly available. Usually, we have to index (annotate) media data, in order to interact or work with them. Indexing is the process of attaching content-based labels to the media. For instance existing literature on video indexing implicitly defines video indexing as the process of extracting the temporal location of a feature and its value from the video data.

Indexing is generally done manually. But the

number of videos available grows and the new applications demand finer grain access to video, thus automation of the indexing process becomes essential. Presently it is reasonable to say that we can extract automatically the following characteristics classified according to the media they are coming from.

**Visual Spatial Content.** We can easily extract a representative image from the visual stream (see video segmentation). From this image, we extract color, texture, sketch, shape, objects and their spatial relationships. The *color* feature is typically represented by image histograms and intersection is used to find similar images (invariant to rotation or translation). The *texture* describes the contrast, the uniformity, the coarseness, the roughness, the frequency, the directionality.

In order to obtain these features either statistical techniques are used (autocorrelation, co-occurrence matrix) or spectral techniques as for instance detection of narrow peaks in the spectrum. The *sketch* gives an image containing only the object outlines and it is usually obtained by edge detection, thinning, shrinking. The *shape* describes global features as the circularity, eccentricity and major axis orientation, but also local ones such as for instance point of curvature, corner location, turning angles and algebraic moments.

**Visual Temporal Content.** One important characteristic of the video is its temporal aspect. We can easily extract the following motions. The *camera* motion describes the real (translation and rotation) and factual movement of the camera (zoom in and out). It is usually obtained either by studying the optical flow by dividing the video in several regions [12] or by studying the motion vector [11]. The *object* motion describes the trajectory obtained by tracking one object on the screen [9].

**Audio.** From the audio stream, we can extract the following basic characteristics. The *loudness* is the strength of sound, determined by the amplitude of the sound wave. The *frequency* translates what we perceive as pitch. The *timbre* is the characteristic distinguishing sounds from dif-

ferent sources. We can easily follow a particular timbre, without recognizing them. More sophisticated characteristic can be obtained as for instance speaker tracking and noise, silence and speech segmentation. To *track a speaker* means that we look for the person who speaks. Usually, at the beginning, a type of voice is quickly learnt and then we try to recognize it during the whole audio stream. The most used models are hidden Markov chains. Another interesting feature is to recognize when *noise, music, speech, or silence* is predominant. Different approaches exist such as the use of expert systems [3] or hidden Markov chains.

**Text.** We are able to manipulate very efficiently the text. So, several researches are done in order to coordinate or to extract text from the video. One research direction is to synchronize the written script with the video. Another one is to extract the text by doing a transcription of the audio channel. This is usually done with complex speech recognition techniques on preprocessed data. Another interesting challenge is to extract the written information appearing on the screen. The idea is to locate the text on the screen and then to recognize it. These techniques are a first attempt, but very often the synchronized text is available and can be exploited.

**Video Structure.** In video we can also easily extract information regarding the structure by doing video segmentation and again extracting a representative image (called key-frame).

The purpose of *video segmentation* is to partition the video stream into basic units (shots) in order to facilitate the indexing, the browsing and to give some structure like paragraphs in a text document. We are not only able to find the shots but also the transition between them, as for instance fade in, fade out, dissolve, wipe. The techniques used in the uncompressed domain are based on pixel-wise or histogram comparison [5]. In the compressed domain [15] they are based either on coefficient manipulations as inner product or absolute difference or on the motion vectors.

The *key-frames* are usually extracted to reduce the image processing to one image per shot. The idea is to find a representative frame (containing

characteristic features) from the sequence of video frames in one shot. One simple method consists in extracting the first or the tenth frame [15]. More sophisticated methods look for local minima of motion or significant pauses [13].

#### 4 Multimedia systems and data mining

In current research [10] where data mining and multimedia systems are connected, the data mining process is often only linked to a particular kind of data, either on image or on a database constructed from the multimedia data.

For instance, in [14], the authors propose the mining of multimedia information from large multidimensional database. Their system is based on an OLAP (on line analytical processing) management system where the multimedia data have been stored after a cleaning step. Afterwards, the process of mining is rather a classical one.

Our aim here is to extract knowledge from the whole kind of available data and to bring up a collaborative data mining process where each kind of data will help to extract a global knowledge. In addition, the introduction of fuzzy set theory in this process enhances the understandability of obtained knowledge.

Moreover, given the current state of the art reliable and efficient automatic indexing is only possible for the presented low-level characteristics. But is clear that any intelligent interaction with the multimedia data should be based on a higher level of description.

The current intelligent systems [2] use high level indexing as for instance a predefined term index or even ontological categories. Unfortunately, the high level indexing techniques are based on manual annotation. So, these approaches can only be used for small quantities of new video and do not exploit intelligently the automatic extracted information.

Also the quantity of information extracted by the indexing algorithms can be extremely large. In addition, the characteristics extracted by the automatic techniques are clearly not sharply defined (color, texture) or they are defined with a degree

of truth (camera motion: fade-in, zoom out) or imprecise (30% noise 50% speech).

So, by applying data mining techniques to the raw multimedia data (indexes), we are going to be able to obtain some knowledge at a higher level. In our case, we propose to use fuzzy decision trees in order to obtain rules that summarize and explain the data. An example of application will be to obtain macro-segmentation.

We consider that such an approach is interesting because it is not only a global approach (to be compared with the micro-segmentation techniques). But also because our aim is not only to extract knowledge from the different indices (color, camera motion, etc.), but also to extract it from different media (audio, video, text). In addition, the use of fuzzy set theory enhances the understandability of the obtained knowledge.

## 5 Conclusion

In this paper, we present the possibilities offered by fuzzy inductive learning and the multimedia indexing techniques which show the potential existing in fuzzy multimedia data mining. Our aim is to use the richness of the multimedia information to extract knowledge by means of fuzzy data mining tools that handle numerical values with fuzzy set theory.

## References

- [1] B. Bouchon-Meunier, C. Marsala and M. Ramdani (1997). Learning from Imperfect Data. In *Fuzzy Information Engineering: a Guided Tour of Applications*, D. Dubois, H. Prade and R. R. Yager eds, chapter 8, pp. 139-148, 1997.
- [2] M. Davis (1993). Media streams: An iconic visual language for video annotation. In *IEEE Symposium on Visual Languages*, pp. 196-202. IEEE Computer Society, 1993.
- [3] M. De Santo, G. Percannella, C. Sansone and M. Vento (2001). Classifying Audio Streams of Movies by a Multi-Expert System. In *Proc. of Int. Conf. on Image Analysis and Processing (ICIAP01)*, Palermo, Italy, Sept. 26-28, 2001.
- [4] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth (1996). From Data Mining to Knowledge Discovery in Databases. In *AI Magazine*, 17:3, 1996.
- [5] F. Idris and S. Panchanathan. Review of image and video indexing techniques (1997). In *Journal of Visual Communication and Image Representation*, 8:146-166, 1997.
- [6] P. Joly (1996). Consultation et analyse des documents en image animée numérique. *Thèse de l'Université P. Sabatier - Toulouse*, 1996.
- [7] R. Kosala and H. Blockeel (2000). Web Mining Research: A Survey. In *SIGKDD Explorations*, volume 2, issue 1, pp. 1-15, 2000.
- [8] D. Kraft, G. Bordogna and G. Pasi (1999). Fuzzy Information Retrieval, in *International Handbook of Fuzzy Sets*, J.C. Bezdek, D. Dubois and H. Prade eds., Kluwer Academic Pub., Vol.3, chapter 6, pp. 469-510, 1999.
- [9] S.Y. Lee and H.M. Kao (1993). Video indexing - an approach based on moving object and track. In *SPIE*. 1908:81-92. 1993.
- [10] First Workshop of Multimedia Data Mining (2000). <http://www.cs.ualberta.ca/zaiane/mdm.kdd2000/>, 2000.
- [11] M. Pilu (1997). On using raw MPEG motion vectors to determine global camera motion. *Tech. report of Digital Media Dept. of HP Lab. Bristol*, Aug. 1997.
- [12] G. Sudhir and J.C.M. Lee (1996). Video annotation by motion interpretation using optical flow streams, *Journal of Visual Communication and Image Representation*. 7:354-368. 1996.
- [13] H.H. Yu and W. Wolf (1999). A hierarchical multiresolution video shot transition detection scheme. *Computer Vision and Image Understanding*. 75:196-213. 1999.
- [14] O.R. Zaïane, J. Han, Z.-N. Li, S.H. Chee and J.Y. Chiang (1998). MultiMediaMiner: A System Prototype for Multimedia Data Mining. In *Proc. of the ACM Sigmod, Int. Conf. on Management of Data*, pp. 581-583, 1998.
- [15] H.J. Zhang, C.Y. Low, S.W. Smoliar and J.H. Wu (1995). Video parsing, retrieval and browsing: an integrated and content-based solution. In *Proc. of ACM Multimedia 95 - Electronic Proceedings*. San Francisco, CA, Nov. 1995.