

Television Stream Structuring with Program Guides

Jean-Philippe Poli^{1,2}

¹ LSIS (UMR CNRS 6168)
Université Paul Cezanne
13397 Marseille Cedex, France
jppoli@ina.fr

Jean Carrive²

² Institut National de l'Audiovisuel
Research and Experimentation Department
94366 Bry-sur-Marne Cedex, France
jcarrive@ina.fr

Abstract

We propose in this paper an original approach to the TV stream structuring problem. The goal of our work is to automatically break the TV stream into telecasts and advertisings and to label each telecast with its genre.

One can think the TV stream structuring problem can be solved by an alignment of the program guide on the stream. But our study shows that, in average, only 25% of the telecasts per day are presented in the program guide. Hence, our method consists in improving statistically these program guides in order to reduce the TV stream structuring problem to a simple alignment problem. The improvement consists in adding the missing telecasts.

We present an original system that lays on the modeling of past TV schedules by a Contextual Hidden Markov Model and a regression tree. Interesting results are presented at the end of the paper.

Keywords : Television Stream Structuring, Contextual Hidden Markov Model, Regression Trees

1. Introduction

The French National Audiovisual Institute¹(INA) is in charge of the TV legal deposit: forty channels are then recorded continuously. INA is used to describing each telecast in order to perform efficiently documents retrieval on its huge database. The structuring of the channels' streams is then a necessary preliminary step, because it isolates all the telecasts and all the advertisings. Television stream structuring can be viewed as the computation of a table of content for a TV stream. The video indexing community did not really interest in video stream structuring, but it proposes several solutions for video indexing and structuring[14]. Video structuring is generally based on video and audio features extraction [5] and integration

	TF1	France 2	France 3	M6
Average telecasts number per day	126	140	161	132
Average of telecasts not shown in PG in %	72.83	80.69	80.53	82.97
Minimum of telecasts not shown in PG in %	84.89	71.82	74.80	76.86
Maximum of telecasts not shown in PG in %	79.51	94.94	84.85	87.60

Table 1. Comparison between real schedules and program guides (PG). TF1 and M6 are private channels and France 2 and France 3 are public ones. The study concerns telecasts broadcast from January 1st 2003 to December 31st 2005.

[7]. Good results are obtained but they are really dependant on telecasts genres: for example, to structure a tennis video, authors of [7] use their knowledge about this kind of game. It may be difficult to define rules for each kind of programs and the cost of their computations is too heavy to be processed on so long documents. Researchers interest also in video genre recognition [13]. It is also based on heavy computations for features extraction but it can only separate very different genres like news or fictions. Nevertheless it can be helpful to differentiate various genres of movie like drama, comedy and horror film.

Program guides (from TV magazines or online Electronic Program Guides) provide a structure for TV streams. One can think that they can be aligned on the stream in order to perform the structuration. Even if they are available at least one week before the broadcast, they cannot be used in their rough state. We have studied the differences between program guides and real schedules. A TV schedule is the exact list of telecasts broadcast on a day. The results are presented in table 1. It shows how much TV guides are incomplete and unusable for an alignment. The study precises that these telecasts that are not presented in program guides (75% of the program guide in average) represent approximately 6 hours per day. These telecasts are

¹<http://www.ina.fr>

advertisings, previews, trailers, lotteries, weather forecast, services (traffic) and small magazines (sponsored or not). The rate of unrepresented telecasts vary according to days. However, program guides give a first idea of the structure. We propose to preprocess program guides in order to statistically improve them. The improvement consists in adding the small telecasts and the advertisings that do not appear in program guides. The result of this improvement must drive detectors by telling them what must be detected and approximately where in the stream. This novel approach decreases the computation cost because detections are not performed on each frame of the stream but only locally.

In the next section, we present the system that we have designed. We then describe each of its parts and we finish by presenting some results of the improvement.

2. System Overview

Figure 2 presents an overview of the system. The goal is to find a structure, a table of content for an input TV stream. The main idea of this approach is the reduction of the TV stream structuring problem into a simple alignment problem by improving program guides.

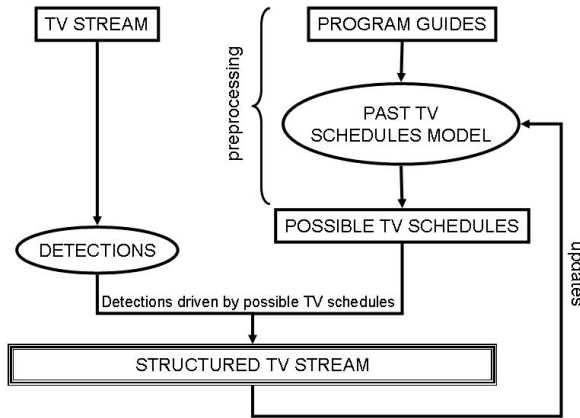


Figure 1. System overview

The program guide in input is combined with past schedules in order to generate all possible schedules for one day. They are generated by both a Contextual Hidden Markov Model (CHMM) and a regression tree. The result can be seen as a tree where each node is a telecast defined by a start hour, a genre and a range of duration given by the regression tree. Each edge is labeled with the transition probabilities given by the CHMM and the tree is explored by choosing the most probable path in the tree. When a node is reached, detections - by automatic detectors that work on the signal - are performed locally from the start hour increased by the minimum duration to the start hour increased by the maximum duration. Detections are used to perform

the alignment. Detectors are chosen in function of the program genre. For example, if the next node is *advertisings*, a commercial detector will be launch.

Detectors can be very general, like a silence detector, or very specific, for instance a channel-specific commercial detector. If they do not detect the end or the beginning of supposed telecast, another path of the tree must be explored. The improvement phase permits to know what to find and where to find the telecasts' boundaries.

3 Improvement phase

In order to statistically improve program guides, a statistical model is required. Markov models[10] are very used for representing sequence of observations. They have been successfully used for video structuring[8]. In order to model TV schedules, we introduced CHMM that are an extension of Hidden Markov Models (HMM) with contextual probabilities. An example of the inadequacy of classical HMM and more details on CHMM can be read in [12].

3.1 Telecasts sequence modeling

3.1.1 Contextual Hidden Markov Models

Definition 1 (Context) A context θ is a set of variables x_1, \dots, x_n with values in continuous or discrete domains, respectively $\{D_1, \dots, D_n\}$. An instance θ_i of this context is an instantiation of each variables x_i :

$$\forall i \in \{1, \dots, n\}, x_i = v_i \text{ with } v_i \in D_i. \quad (1)$$

From this point, we also call θ_i a context. It is possible to update a context θ_i into a context θ_{i+1} with an evolution function.

Definition 2 (Evolution function) Let Θ be the set of all possible instances of a context θ . An evolution function F for θ is defined by:

$$F : \Theta \times D_{p_1} \times \dots \times D_{p_m} \rightarrow \Theta \\ \theta_i, p_1, \dots, p_m \rightarrow \theta_{i+1} \quad (2)$$

where D_{p_i} is the domain of the external parameter p_i .

We can now introduce Contextual Hidden Markov Models (CHMM) which are basically a Markov model where the probabilities are not only depending on the previous state but also on a context. This context is updated every time a state of the model is reached.

Definition 3 (Contextual hidden Markov models) A contextual hidden Markov model is totally defined by $\langle S, \Sigma, \Theta, F, \pi_\theta, A_\theta, B_\theta \rangle$, where:

- S is a state space with n items and s_i denotes the i^{th} state in the state sequence,

- Σ is an alphabet with m items and ϵ_j denotes the j^{th} observed symbol,
- Θ is the set of all instances of the context θ ,
- F denotes the evolution function for instances of θ ,
- π_θ is a parametrized stochastic vector and its i^{th} coordinate represents the probability that the state sequence begins with the state i . π_i is a function of θ which represents the initial distribution in the context θ :

$$\forall i \in \{1, \dots, n\}, \pi_i(\theta_1) = P(s_1 = i | \theta_1), \quad (3)$$

- A is a stochastic matrix $n \times n$ where a_{ij} stands for the probability that the state i is followed by state j in the state sequence. Each a_{ij} is a function of θ :

$$\forall k, t \in \mathbb{N}, \forall i, j \in \{1, \dots, n\}, \quad (4)$$

$$a_{ij}(\theta_k) = P(s_{t+1} = j | s_t, \theta_k),$$

- B is a stochastic matrix $n \times m$ where b_{ik} represents the probability of observing the symbol k from state i :

$$\forall k, t \in \mathbb{N}, \forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, m\} \quad (5)$$

$$b_{ij}(\theta_k) = P(\epsilon_t = j | s_t, \theta_k).$$

Probabilities in a contextual semi-Markov model depend only on the current context (not the previous or following ones). The observed symbols are all independent and transition probabilities depend only on the previous state.

The context permits to resolve certain ambiguities in the transitions and eliminates impossible transitions in a particular context. We can expand the context to seasons and vacations to be closer to the reality. But presently, we only regard broadcast times and days.

3.1.2 Application to TV schedules modeling

In order to represent the TV schedules, we chose to attribute at each state of the CHMM a telecast genre. We chose a continuous distribution for the emission probabilities : this means that observations are not discrete in our case. When we are on a state of our CHMM, for example the state representing magazines, we have a continuous distribution over its possible durations.

The context θ for our model can be a variable Hr that represents the hour of beginning of a telecast by an integer in the range $\{0, \dots, 86399\}$, and a variable Day that represents the broadcast day of week with an integer in the range $\{0, \dots, 6\}$:

$$\theta = \{Hr, Day\}$$

$$\text{and } D_{Hr} = \{0, \dots, 86399\}, D_{Day} = \{0, \dots, 6\}.$$

The evolution function F simply consists in an addition of the length of a telecast to the previous context. Let see now how the probability of a schedule can be evaluated. Let $\langle Monday, 6 : 30, Magazine, 10min \rangle$ denotes a magazine that starts on Monday at 6:30 a.m. and that lasts 10 minutes. Let M be a CHMM. Then, the probability of the schedule S such as:

$$S = \langle Monday, 6 : 30, Magazine, 10min \rangle$$

$$\langle Monday, 6 : 40, IP(inter - programs), 3min \rangle \quad (6)$$

$$\langle Monday, 6 : 43, News, 20min \rangle$$

can be written:

$$P(S|M) = P(magazine|\{monday, 23400\})$$

$$\times P(d = 10min|magazine, \{monday, 23400\})$$

$$\times P(IP|\{monday, 24000\}, magazine)$$

$$\times P(d = 3min|IP, \{monday, 24000\})$$

$$\times P(news|\{monday, 24180\}, IP)$$

$$\times P(d = 20min|news, \{monday, 24180\}). \quad (7)$$

As shown in equation 7, it is necessary to estimate the probability of a particular duration. We present in the next section our method to predict durations of a particular telecast.

3.2 Duration probability estimation

3.2.1 Regression trees

Regression trees [1] are tools for predicting continuous variables or categorical variables from a set of mixed continuous or categorical factor effects. Regression trees are used to predict continuous values from one or more predictor variables. Their prediction are based on few logical *if-then* conditions. A regression tree is a tree where each decision node in it contains a test on some predictor variables' value. The leaves of the tree contain the predicted forecast values. Regression trees are built through a recursive partitioning. This iterative process consists in splitting the data into partitions (generally two partitions), and then splitting them up further on each of the branches. The chosen test is the one that satisfies a user-defined criteria.

3.2.2 Application to television schedules modeling

We use a regression tree in order to resolve two different problems. Firstly, we use it to predict a range of durations for a telecast from its context (i.e. broadcast days and hours, previous telecast). It is very useful to know that between the minimum duration and the maximum duration a telecast transition may occur in order to only look for it in this temporal window. But this problem is directly resolved by regression trees. Secondly we want to deduce a probability from a leaf of the regression tree.

We represent the distribution of the durations on a leaf with the asymmetric gaussian presented in [6]. Let μ and σ be respectively the mean value and $|Min(Duration) - Max(Duration)|$. Then the probability of a given duration d is given by:

$$A(d, \mu, \sigma^2, r) = \frac{2}{\sqrt{2\pi}} \frac{1}{\sigma(r+1)} \begin{cases} e^{-\frac{(d-\mu)^2}{2\sigma^2}} & \text{if } d > \mu \\ e^{-\frac{(d-\mu)^2}{2r^2\sigma^2}} & \text{otherwise} \end{cases} \quad (8)$$

$$\text{where } r = \frac{|\mu - \min(Duration)|}{|\mu - \max(Duration)|}.$$

3.3 Combining program guides and model's predictions

We have introduced a model that can represent TV schedules. More recent informations about the stream are provided by TV guides, which are delivered at least one week before the broadcast. In the better case, the program guide is included in the schedules predicted by the model: there is no need to revise the schedule. In another case, the program guide is in contradiction with the predicted schedules: then they need to be combined. In the worst case, the program guide does not match with what has been broadcast (a special and unforeseeable event occurs): the system cannot work on special streams and the structuring must be done manually.

The difficulty of combining both the predictions and the program guide is the telecast matching. A telecast that appears in the prediction must fit a telecast in the program guide while they do not have the same duration and the same start hour. To perform this matching, we use an elastic partial matching method [9]. The proposed algorithm resolves the best matching subsequence problem by finding a cheapest path in a directed acyclic graph obtained from the two input sequences of values. It can also be used to compute the optimal scale and translation of time series values. The algorithm needs a distance to compare the values; in their case, they use the euclidean distance between two real values. We have used the following measure d between two telecasts E_1 and E_2 : $d(E_1, E_2)$ returns ∞ if E_1 and E_2 have not the same genre, and it returns $|E_1.Start - E_2.Start| + |E_1.Duration - E_2.Duration|$ otherwise.

In order to make the combination, we consider that the first telecast of both the program guide and the prediction is synchronized with the real start hour of the telecast. The method consists then in predicting telecasts from a telecast of the program guide to the next one. If we consider the predicted schedules as a graph, it maps with browsing the graph in depth-first order until a telecast matches with the next telecast of the program guide. We introduced a threshold Δ which specifies the maximal delay between a telecast from the prediction and a telecast from the program genre. If the algorithm passes this delay, we consider a matching telecast will not be found. We then add the unmatched telecast from the program guide to the graph of predictions and the CHMM is reinitialized with the new context.

The algorithm selects the possible paths in the prediction tree regarding the program guide. In order to decrease the combinatory aspect of the algorithm, two heuristics are used.

Heuristic 1 : Pruning the impossible branches. We made a list of telecast genres that must appear in a program guide. For example, movies and TV shows always appear

in a program guide, contrary to weather forecast, short magazines which can be omitted. If a path between two successive telecasts in the program guide passes by a telecast whose genre always appears in program guides, then the path can be pruned.

Heuristic 2 : Merging matching telecasts. Several paths can lead from one telecast of the program guide to the following one. Thus, there are several matching telecasts which differ from start hours and sometimes from durations. However, they represent the same node and then can be merged.

4 Alignment phase

The next phase of the TV stream structuring is the alignment of the improved program guides on the stream itself. This phase requires detectors in order to find locally in the stream the end or the beginning of each program contained in the improved program guides. We are still testing and looking for novel solutions for this phase.

In [15], authors use monochromatic frames detection conjointly with silence detection in order to find breaks in the stream. This method suffers from the number of false alarms that occur inside a telecast. Detection as simple as this one may not be sufficient. The author of [11] proposes jingle recognition that can be useful for TV themes recognition: final credits of TV series can be detected with this method.

The main difficulty of this phase is the commercials and trailers detection. This two genres are really numerous on a day. Several solutions have been proposed in literature but they are impractical in France because of the preceding and the following jingles and French regulations. They are based on blank frames detection[3] or multimodal features [4]. Another solution consists in detecting them as duplicate sequences[2]; but trailers are not always broadcast several times a day or a week. In France, commercials and trailers are preceded and followed by special jingles that vary according to days, hours and special events. We are working on commercials and trailers detection by automatically finding invariant features in their jingles (like a logo or a sound).

5 Results

The statistical improvement of program guides has been implemented. We present in this section some of the results we obtained.

We consider 36 different genres of telecasts. A broadcast day is composed by 120 telecasts in average (table 1). There are hence $36^{120} \approx 5.7 \times 10^{186}$ possible schedules. The model decreases the number of possible schedules by

deleting impossible successions (for instance a day composed by 120 telecasts of the same genre). In order to test the model, we trained it on telecasts broadcast on France 2 in 2004 (it represents more than 50000 telecasts) and we tested the model on one week in 2005. Without the application heuristics, we had approximately 150000 possible paths that reach the 10th telecast on Friday may 2th. With heuristics 1 and 2, we have only 7 possible paths. Heuristics really speed up the prediction but some paths would be kept.

For the regression tree, we fixed $\omega = \nu = 300$. That means the minimum width of a temporal window is 300 seconds. We have 97% of good predictions. Good predictions are durations that are between the minimum and maximum values given by the leaf of the regression tree.

The CHMM can represent 83% of the days in 2005. The others present special events. We fixed $\Delta = 1800$, i.e. a delay of 30 minutes between the start hour in the program guide and the real schedule is authorized. The improvement of 7 schedules from a program guide gives from 3 to 6 possible schedules. Only one of them is correct if we compare them to the ground truth. With all heuristics, when at least one path exists between two consecutive telecasts, only few nanoseconds are necessary. Otherwise, if there is no path and if a telecast from program guide must be added, it takes up to 20 seconds in average. For the prediction of a TV schedule, it takes less than 2 minutes in average.

Results could surely be ameliorated by cleaning up the training and the testing sets. In fact, special events like the Pope's death and Olympic Games have not been removed and change certain probabilities.

6 Conclusion

We present in this article a novel approach for television video structuring. The main idea is the use of program guides and their improvement with knowledge from the past schedules in order to avoid heavy computation with features extraction, detections and recognitions.

The improvement is performed with a Contextual Hidden Markov Model that gives all possible schedules, according to past experience, for a particular day. A regression tree is used in order to predict telecasts' durations range. This creates temporal windows in which a telecast may end or begin.

Results of the improving part of the system have been presented. The next step is to drive detectors while browsing the tree of possible schedules.

7 Acknowledgement

The research work leading to this paper has been partially supported by the European Commission under the IST

research network of excellence K-SPACE.

References

- [1] L. Breiman, J. Friedman, R. Olshen, and C. Stone". "classification and regression trees". Technical report, "Wadsworth International, Monterey, CA, USA", 2004.
- [2] P. Duygulu, M.-Y. Chen, and A. Hauptmann. Comparison and combination of two novel commercial detection methods. In *The 2004 International Conference on Multimedia and Expo (ICME'04)*, June 2004.
- [3] D. S. et al. *Automatic TV advertisement detection from mpeg bitstream*, volume 35, pages 2–15. 2002.
- [4] S. M. et al. Audio and video processing for automatic tv advertisement detection. In *Proceedings of ISSC 2001*, 2001.
- [5] D. Gatica-Perez, M. Sun, and A. Loui. Probabilistic home video structuring: Feature selection and performance evaluation. In *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2002.
- [6] T. Kato, S. Omachi, and H. Aso. Asymmetric gaussian and its application to pattern recognition. In *Lecture Notes in Computer Science (Joint IAPR International Workshops SSPR 2002 and SPR 2002)*, volume 2396, pages 405–413, 2002.
- [7] E. Kijak, L. Oisel, and P. Gros. Audiovisual integration for tennis broadcast structuring. In *International Workshop on (CBMI'03)*, 2003.
- [8] E. Kijak, L. Oisel, and P. Gros. Hierarchical structure analysis of sport videos using hmms. In *IEEE Int. Conf. on Image Processing, ICIP'03*, volume 2, pages 1025–1028. IEEE Press, 2003.
- [9] L. J. Latecki, V. Megalooikonomou, Q. Wang, R. Lakaemper, C. A. Ratanamahatana, and E. Keogh. Partial elastic matching of time series. *icdm*, 0:701–704, 2005.
- [10] J. Norris. *Markov chains*. Cambridge series in statistical and probabilistic Mathematics, 1997.
- [11] J. Pinquier. *Primary audio features for audiovisual structuring*. PhD thesis, Université Paul Sabatier (Toulouse III), 2004.
- [12] J.-P. Poli and J. Carrière. Improving program guides for reducing tv stream structuring problem to a simple alignment problem. In *Proceedings of CIMCA 2006*, November 2006. To appear.
- [13] M. Roach, J. Mason, and M. Pawlewski". "video genre classification using dynamics". In *"IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2001"*, volume "3", pages 1557–1560, "2001".
- [14] C. G. Snoek and M. Worring. Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications*, 25(1):5–35, 2005.
- [15] P. G. Xavier Naturel, Guillaume Gravier. étiquetage automatique de programmes de télévision. In *Proceedings of CORESA'05*, 2005.