

Apprentissage de structure dans les réseaux bayésiens pour la détection d'événements vidéo

Siwar Baghdadi¹, Claire-Hélène Demarty¹, Guillaume Gravier², et Patrick Gros³

¹ Thomson R&D France,
1 av Belle Fontaine-CS 17616,
35576 Cesson-Sévigné. France
{siwar.baghdadi, claire-helene.demarty}@thomson.net

² CNRS, IRISA
Campus de Beaulieu
35042 Rennes Cedex, France.
{guillaume.gravier}@irisa.fr

³ INRIA
centre de Rennes - Bretagne Atlantique
Campus de Beaulieu, 35042
Rennes, France
{patrick.gros}@inria.fr

Résumé Dans cet article, nous proposons un système de détection d'événements basés sur les réseaux bayésiens. Ce système repose sur l'apprentissage automatique de la structure du réseau bayésien à partir de données d'apprentissage. Nous comparons notre approche à une approche naïve qui suppose que les attributs du système sont indépendants conditionnellement à l'événement. Afin de valider notre approche, nous considérons la détection d'*Actions* dans un match de football. Les résultats expérimentaux effectués sur une base de quatre matchs montrent une amélioration des performances de classification par rapport à la méthode naïve utilisée classiquement dans la littérature.

Mots clés Réseaux bayésiens, apprentissage de structure, classification, indexation vidéo.

1 Introduction

Plusieurs travaux proposant de faire l'indexation vidéo dans le cadre des réseaux bayésiens sont apparus récemment dans la littérature. Dans [1], les auteurs utilisent ainsi un réseau bayésien pour extraire les moments intéressants dans des vidéos de Formule 1. Wang et al [2] ont quant à eux mis en place un système utilisant les réseaux bayésiens pour la détection d'événements dans les vidéos de sport. Ces deux approches utilisent une structure du réseau bayésien construite à la main, à partir de connaissances sur le domaine traité. Les connaissances sur les différentes relations qui existent entre les variables du problème ne sont cependant pas toujours disponibles. A défaut, la structure est fixée, toujours manuellement, mais en supposant généralement l'indépendance entre les attributs conditionnellement à l'événement recherché.

Il est toutefois également possible d'exploiter les données d'apprentissage pour construire automatiquement la structure du réseau bayésien. Dans de précédents travaux [3], nous proposons d'utiliser l'apprentissage de structure pour construire automatiquement le modèle dans le but de faire la détection de publicité dans un flux TV. L'approche utilisée dans [3], qui apprend un réseau bayésien non contraint, ne distingue cependant pas

le nœud de classification des autres nœuds. Cette méthode n'est pas optimale dans le cas de la classification d'événements rares tels que les *Actions* ou les *Buts* dans les matchs de football. Nos expériences montrent en effet qu'une recherche contrainte, où un rôle particulier est donné au nœud de classification, s'avère plus efficace. Après une description du concept des réseaux bayésiens, nous présentons les différentes méthodes d'apprentissage de structure que nous utilisons. Nous poursuivons par la présentation des résultats de ce type d'apprentissage sur l'application que nous avons envisagée, avant de conclure.

2 Les réseaux bayésiens

La théorie des réseaux bayésiens résulte d'une fusion entre la théorie des probabilités et la théorie des graphes [4]. On définit classiquement un réseau bayésien comme un graphe acyclique dirigé. Il est formé d'un ensemble de variables et d'un ensemble d'arcs entre les variables. Chaque variable correspond à un nœud du réseau. A chaque variable X_i ayant pour parents l'ensemble : $pa(X_i)$, on associe une probabilité conditionnelle $p(X_i|pa(X_i))$.

Dans les réseaux bayésiens, la probabilité jointe du modèle s'écrit ainsi :

$$P(X_1, \dots, X_n) = \prod_{1..n} (P(X_i|pa(X_i))) \quad (1)$$

Un réseau bayésien possède deux niveaux de paramètres : des paramètres quantitatifs qui sont les probabilités conditionnelles associées à chaque nœud, $p(X_i|pa(X_i))$, et des paramètres qualitatifs qui sont les arcs entre les différents nœuds. L'ensemble de ces arcs forme la structure du réseau.

Deux types d'apprentissage sont disponibles dans les réseaux bayésiens. Le premier type, largement utilisé dans la littérature, est un apprentissage de paramètres. Généralement, cet apprentissage se fait avec la méthode du maximum de vraisemblance. Le deuxième type est l'apprentissage de structure.

Dans [3], les auteurs proposent d'utiliser l'algorithme K2 [5], pour réaliser l'apprentissage de structure d'un réseau bayésien non contraint. Ils obtiennent ainsi une structure générique pour la détection de publicité. Cette méthode d'apprentissage de structure utilise le score BIC pour évaluer les structures. Il peut se décomposer au niveau de chaque nœud X_i sous la forme de l'équation 2.

$$score_{BIC}(X_i, p(X_i)) = \log(P(X_i|pa(X_i))) - \frac{1}{2} \cdot Dim(X_i, \mathcal{G}) \cdot N \quad (2)$$

où N est le nombre d'exemples dans la base de données, et $Dim(X_i, \mathcal{G})$ est le nombre de paramètres nécessaires pour décrire l'information au niveau du nœud dans la structure \mathcal{G} .

Le score BIC est ainsi composé de deux termes, un premier terme qui tient compte de la vraisemblance des données par rapport au modèle, c'est donc un terme d'attache aux données. Le second terme tient compte de la complexité de la structure. Cette formulation du score ne met cependant aucunement en avant le nœud de classification. Dans ce cadre, la méthode d'apprentissage de structure basée sur le score *BIC* de l'équation 2 peut résulter en une structure qui simulera correctement les données sans toutefois être optimale pour notre tâche de classification. Dans le paragraphe suivant nous proposons donc de revoir cette méthode de façon à l'adapter à la classification d'événements.

3 Utilisation de l'apprentissage de structure pour la classification

Notre objectif principal est la détection d'événements dans les vidéos. Nous étudions dans cette partie l'utilisation de l'apprentissage de structure pour automatiser cette tâche de détection d'événements. Le réseau bayésien le plus connu dans la littérature est le réseau bayésien naïf, connu aussi sous le nom de classifieur bayésien. Dans ce type de réseau, les attributs X_1, \dots, X_{n-1} sont supposés indépendants conditionnellement à la classe X_c . Les nœuds attributs ne possèdent qu'un seul parent, c'est le nœud de classification. Cette hypothèse entraîne la simplification de la loi jointe de l'équation 1 sous la forme de l'équation 3. De tels réseaux ont été largement utilisés dans la littérature pour la classification [6]. Ils se caractérisent, en effet, par la rapidité des opérations d'apprentissage et d'inférence.

$$P(X_c, X_1, \dots, X_{n-1}) = P(X_c) \cdot \prod_{i=1}^{n-1} (P(X_i/X_c)) \quad (3)$$

Toutefois, dans les réseaux bayésiens naïfs, aucune corrélation entre les attributs n'est prise en compte. Toutes les caractéristiques contribuent à la classification de la même manière. Le nœud de classification profite de l'information donnée par chaque attribut indépendamment de l'information donnée par les autres caractéristiques. Ceci peut ne pas être optimal pour la tâche de classification. Nous proposons d'enrichir la structure du réseau bayésien naïf pour tenir compte des corrélations qui existent entre les différents attributs.

Dans [7], les auteurs ont proposé l'approche TAN pour enrichir la structure du réseau bayésien. Cette approche utilise une structure en arbre afin de faire la classification. La structure en arbre présente l'avantage d'avoir une complexité faible; elle évite donc les problèmes de sur-apprentissage. Toutefois restreindre le nombre de parents autres que le nœud de classification à exactement un parent pour chaque nœud, est une contrainte forte. La structure ainsi obtenue ne permet pas de représenter le cas où une variable est corrélée avec plusieurs autres variables. Elle ne permet pas non plus de représenter le cas où une variable est conditionnellement indépendante de toutes les autres variables par rapport au nœud de classification. Dans ce cas, le nœud représentant cette variable n'a besoin que du nœud de classification comme parent. L'ajout d'un autre parent ne fait qu'augmenter inutilement la complexité et le nombre de paramètres du réseau.

Pour ces raisons, nous utilisons l'algorithme K2 pour enrichir la structure du réseau bayésien naïf. Ce choix nous permet de ne plus nous restreindre à une structure d'arbre mais d'avoir une structure plus générique. Nous avons également modifié le score BIC de l'algorithme K2, pour tenir compte du fait que chaque nœud attribut doit avoir comme parent le nœud de classification. Le score *BIC* modifié au niveau de chaque nœud X_i s'écrit alors :

$$score_{BIC}^m(X_i, p(X_i)) = \log(P(X_i/pa(X_i), X_c) - \lambda \cdot Dim(X_i, \mathcal{G}) \cdot N) \quad (4)$$

A l'image du score *BIC*, le score *BIC* modifié est composé de deux termes. Un premier terme permettant la maximisation de la vraisemblance; et un second terme permettant de tenir compte de la complexité du réseau construit. La variable λ permet une pondération

entre l'influence de ces deux termes. Ainsi, plus λ est grand, plus la complexité de la structure aura de poids dans le calcul du score, et plus les structures obtenues seront simples.

4 Résultats expérimentaux

4.1 Protocole expérimental

Afin de tester notre approche, nous avons pris le cadre de la détection d'Actions dans un match de football. Une *Action* est un moment du match où un joueur tente de marquer un but. D'un point de vue vidéo, cela se traduit par un moment du jeu généralement au niveau de la zone de but de l'une des deux équipes, avec une acclamation de la foule et une augmentation du niveau d'excitation du présentateur. Cet instant est aussi généralement suivi par des plans de ralenti.

Les attributs que nous utilisons sont des attributs extraits des signaux audio et vidéo : *niveau sonore de la foule, plage de jeu/non jeu, type de plan (large ou pas), transition, présence de visage, couleur verte, logo du ralenti, position sur le terrain*. Puisqu'une *Action* influence également les plans suivants dans la vidéo nous rajoutons à notre liste d'attributs les attributs des cinq plans suivants. Nous totalisons ainsi 40 variables pour les attributs, et une variable pour la classe *Action*. Notre base de données est constituée de quatre matchs issus de la coupe du monde 2006, ce qui correspond à 109 *Actions* et 6300 plans autres. Notre base de données n'étant pas très grande, pour la phase de test, nous avons choisi d'utiliser un processus de cross-validation. Nous présentons nos résultats sous forme de courbe de Précision/Rappel.

4.2 Résultats

Nous proposons dans la figure 1 une comparaison entre l'approche utilisant un réseau bayésien non contraint, l'approche utilisant les réseaux bayésiens naïfs et l'approche enrichissant les réseaux naïfs par une structure générique telle que proposée dans le paragraphe précédent. Il apparaît clairement que les résultats du réseau bayésien non contraint ne sont pas satisfaisants en terme de classification. Ces résultats sont même de moindre qualité que ceux donnés par un réseau bayésien naïf, classiquement utilisé dans les tâches de classification. Ces résultats peuvent être expliqués par le fait que cette approche ne donne aucune position particulière au nœud de classification. Elle cherche à maximiser la vraisemblance du modèle par rapport aux données, sans tenir compte du fait que notre but principal est la classification. Dans le cas où l'on dispose d'un nombre important d'attributs, le terme tenant compte de la classification est en effet noyé par la vraisemblance des attributs. L'approche que nous avons proposée suppose, quant à elle, que le nœud de classification est connecté à tous les attributs. Elle permet donc au nœud de classification de profiter de l'information de tous les attributs.

Sur la figure 1, il apparaît aussi que notre approche donne de meilleurs résultats que la structure du réseau bayésien naïf, pour lequel on suppose une indépendance conditionnelle

entre les attributs par rapport au nœud de classification. Ce résultat montre l'importance de tenir compte des connexions entre les attributs pour augmenter le pouvoir de classification. Dans la figure 2, nous comparons notre approche à l'approche TAN qui augmente

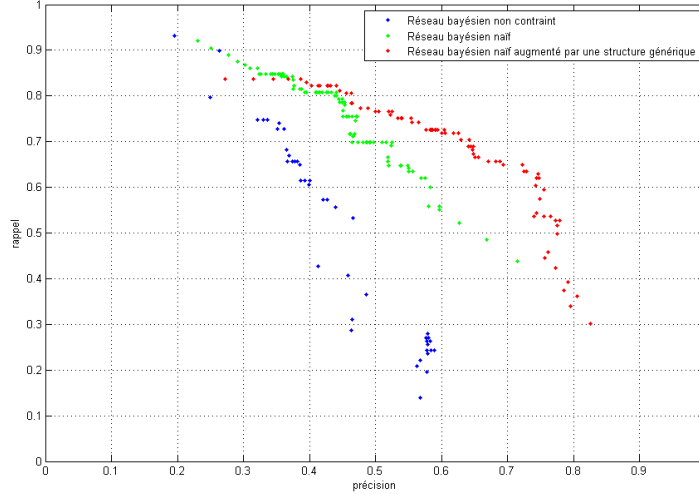


Fig. 1. Comparaison entre les résultats du réseau bayésien naïf, du réseau bayésien non contraint et du réseau bayésien naïf augmenté par une structure générique.

le réseau bayésien par un arbre, présenté dans [7]. La différence entre les deux approches réside dans le fait que pour l'approche TAN, la structure du réseau bayésien naïf est augmentée par une structure d'arbre, alors que notre approche enrichit la structure du réseau naïf par une structure générique. Nous remarquons que l'enrichissement par une structure générique donne de meilleurs résultats de classification. En effet se restreindre à une structure en arbre ne garantit pas des résultats de classification optimaux. Cela permet en outre de tenir compte des corrélations existant entre les attributs.

5 Conclusion

Les tests expérimentaux que nous avons effectués montrent que l'apprentissage de structure des réseaux bayésiens améliore les performances de classification d'événements dans les données vidéo par rapport au réseau bayésien naïf classiquement utilisé dans la littérature. On peut donc conclure que l'apprentissage de structure constitue un outil efficace pour automatiser la tâche de détection d'événements dans la vidéo.

Plus précisément, nous avons aussi montré que l'apprentissage de structure doit gérer le nœud de classification différemment des nœuds attributs. Nous avons en effet prouvé dans notre application que les approches de type réseau bayésien naïf enrichi améliorent les performances de classification du système. De plus, l'utilisation d'une structure générique

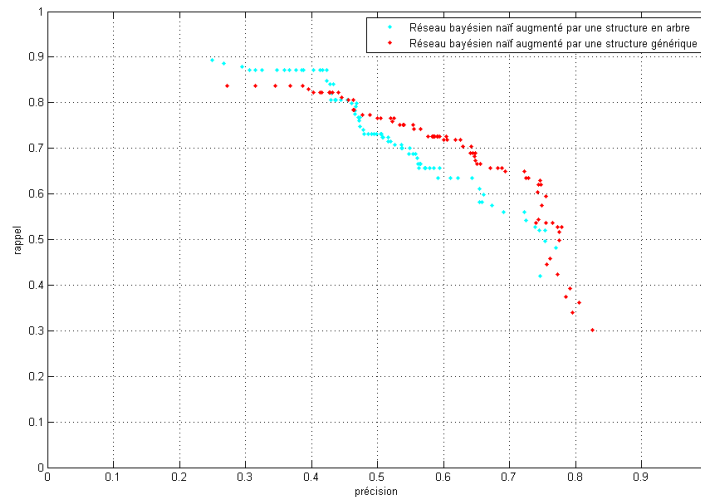


Fig. 2. Comparaison entre les résultats du réseau bayésien augmenté par une structure d’arbre et du réseau bayésien augmenté par une structure générique.

pour enrichir le réseau naïf améliore encore les résultats de classification, par rapport à une approche qui utilise une structure d’arbre.

Dans le futur, nous envisageons d’étudier l’effet de la sélection d’attributs sur les approches telles que les approches de réseaux bayésiens naïfs enrichis. Ce type d’approches, tout comme les réseaux bayésiens naïfs, prend en effet en compte tous les attributs, y compris ceux qui ne sont pas utiles pour la classification.

Références

1. M. Petkovic, V. Mihajlovic, W. Jonker, and S. Djordjevic-Kajan. Multi-modal extraction of highlights from TV formula 1 programs. In *Proc. IEEE ICME*, pages 817–820, 2002.
2. F. Wang, Y. Ma, H. Zhang, and J. Li. A generic framework for semantic sports video analysis using dynamic bayesian networks. In *Proc. IMMC*, pages 115–122, 2005.
3. S. Baghdadi, G. Gravier, C.H. Demarty, and P. Gros. Structure learning in a Bayesian network-based video indexing. In *Proc. IEEE ICME*, pages 677–680, 2008.
4. F.V Jensen. *Bayesian Networks and Decision Graphs*. Springer, 2001.
5. G. F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, pages 309–347, 1992.
6. P. Langley, W. Iba, and K. Thompson. An analysis of Bayesian classifiers. *Proc. NCAI*, pages 223–228, 1992.
7. N. Friedman, D. Geiger, and M. Goldszmid. Bayesian network classifiers. *Machine Learning*, 29(2) :131–163, 1997.