

Application des arbres de décision flous à la reconnaissance des BVPs

M.A. Chikh

Laboratoire de GBM
Université de Tlemcen Algerie
c_bamine@yahoo.fr

P.Y Glorennec

Departement d'informatique
INSA Rennes France
glorenne@irisa.fr

Résumé Cet article présente un modèle de classification explicite des battements ventriculaires prématurés(BVPs) à partir du signal physiologique électrocardiogramme(ECG). Il s'agit de proposer au médecin un ensemble de cas et de scénarios sur les données du patient pour l'aider à établir un diagnostic. Dans ce but une approche de la logique floue, arbre de décision flou(ADDF),a été proposé. Cette technique repose notamment sur les principes de structuration, de complémentarité entre description explicite et discriminante des classes ainsi que sur la notion d'interprétabilité des résultats. Les performances du modèle a été testé sur la base de données MIT-BIH(Massachusetts Institut of Technology - Beth Israel Hospital)

Mots clés : Battements BVPs, ADDF, base de données MIT-BIH .

1 INTRODUCTION

Actuellement l'électrocardiographie est une technique relativement peu coûteuse permettant, à l'aide d'un examen indolore et sans danger, de surveiller l'appareil cardio-vasculaire, notamment pour la détection des troubles du rythme, en étudiant la forme, la fréquence et l'évolution des signaux électrocardiogrammes (ECGs). L'analyse du rythme figure parmi les applications les plus prometteuses de l'ingénierie biomédicale : elle consiste, pour l'essentiel, à détecter les anomalies éventuelles, à identifier leur nature et à préciser un traitement approprié, d'où l'intérêt de développer un outil automatisé d'aide au diagnostic des pathologies cardiaques, fondé sur le traitement systématique du signal ECG(Figure 1). La prise en charge des battements ventriculaires prématurés 'BVPs', lorsque ceux-ci apparaissent en grand nombre (initiation d'une tachycardie ventriculaire), a beaucoup évolué au cours de la dernière décennie. D'une part, l'accroissement des connaissances sur la physiopathologie des arythmies ventriculaires et l'amélioration des techniques d'exploration électrophysiologique permettent d'identifier les patients à risque de mort subite rythmique et de prendre des mesures préventives comme l'implantation d'un défibrillateur cardiaque.

Le complexe QRS reflète l'activité ventriculaire du cœur. Sa morphologie peut changer lorsque sa commande ou la conduction du front de dépolarisation ou de repolarisation devient anormale. Par exemple,

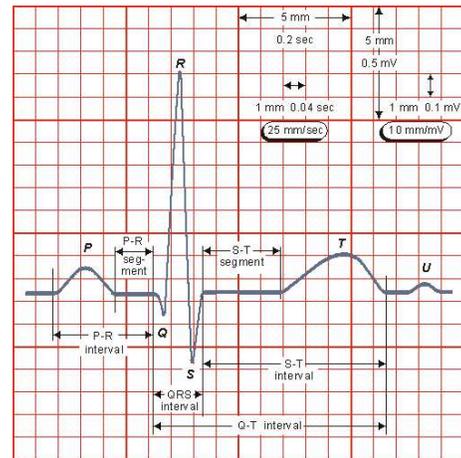


FIG. 1 – Ondes et intervalles de l'électrocardiogramme

un battement ventriculaire prématuré 'BVP' conduit généralement à un élargissement anormal du complexe QRS. La classification du complexe QRS est donc absolument indispensable lors de l'élaboration d'un système de reconnaissance de troubles arythmiques. Il convient de rappeler que de nombreuses approches citées dans la littérature, sous le terme de classification d'arythmies cardiaques ne sont en fait que de simples classificateurs du complexe QRS (Suzuki 1995), (Chow 1995), (Lagerholm 2000).

Le problème traité dans ce papier part du fait que la majorité des systèmes de reconnaissance et de classification des anomalies cardiaques, développés au cours des ces dernières années, présentent des résultats non explicites (Silipo 1999a), (Carrault 2003). Dans le cadre de notre travail, nous proposons une méthode de reconnaissance des arythmies cardiaques basée sur une technique appartenant au domaine de l'intelligence artificielle, à savoir les arbres de décision flous (ADDFs)(Le Guillou 2004). Ces techniques permettent d'identifier la classe de l'anomalie cardiaque tout en présentant une interprétation du résultat obtenu qui est proche de celle du médecin expert.

2 Base d'exemples

L'apprentissage doit être réalisé avec un ensemble d'apprentissage et un ensemble de validation. Ce der-

nier est indispensable pour éviter le sur-apprentissage. Il est utilisé pour tester les capacités de généralisation du modèle de classifieur créé à partir des données d'apprentissage. Nous avons utilisé 15 enregistrements de la base de données MIT-BIH, en choisissant pour chaque enregistrement 30 battements ventriculaires prématurés 'Bvp' et 30 battements normaux (Nm) pour préparer respectivement une base d'apprentissage et une base de validation (chacune des bases est composée de 900 battements). Nous avons écarté les deux enregistrements 105 et 114, qui ne contiennent pas un nombre suffisant de 'Bvp' (voir tableau 1).

TAB. 1 – Les enregistrements choisis de la base de données MIT-BIH

Enregistrements	battements 'Nm'	battements 'BVP'
105	2549	41
106	1498	519
114	1810	44
116	2273	108
119	1535	442
200	1737	820
203	2494	439
205	2557	71
208	1578	985
210	2413	183
213	2624	220
215	3181	164
219	2071	64
221	2022	393
223	2017	473
228	1696	362
233	2220	824

3 Méthode utilisée

3.1 Caractéristiques principales d'un ADDF

Les arbres de décision (Rasoul 1991), (Chidanand 1997), (Sreerama 1998), (Shu 2004) sont parmi les méthodes de classification les plus anciennes et les plus intuitives. Leur modélisation repose sur une structuration précise des connaissances sous la forme d'un graphe orienté (un arbre). Cette organisation les rend particulièrement simples à utiliser et à interpréter. C'est pourquoi cette approche a connu un vif succès dans différents domaines comme l'aide au diagnostic médical (Giesen 1996), (Antonia 2003), (Mia 2003), (Sotiris 2004).

Les arbres de décision flous généralisent les arbres de décision classiques et sont beaucoup mieux adaptés pour le traitement de données numériques continues. Le parcours de la racine à une feuille constitue une règle floue. Lors de la présentation d'un exemple, différentes règles sont activées et un mécanisme d'inférence permet de composer ces règles pour obtenir la réponse de l'arbre.

Les arbres de décision ou de régression¹ flous permettent de manipuler des variables continues sans introduire des seuils arbitraires. Dans un arbre de décision booléen, des tests en tout-ou-rien sont réalisés à chaque nœud.

Dans un ADDF la branche forme la partie prémisses, tandis que la valeur associée à une feuille forme la conclusion. Soit c_N cette conclusion et $\alpha_N(\mathbf{x}_i)$ la valeur de vérité de la règle correspondante pour un vecteur d'entrée $\mathbf{x}_i = (x_{i1}, \dots, x_{iM})$ donné. Chaque attribut, x_{ij} , ayant des modalités floues, avec

$\mu_{A_j}(x_{ij})$ est le degré d'appartenance de x_{ij} au sous-ensemble flou A_j . On appelle aussi

$\alpha_N(\mathbf{x}_i)$ le degré d'appartenance du vecteur \mathbf{x}_i au nœud N calculé de la racine au nœud.

– $\alpha_N(\mathbf{x}_i)$ est calculée à partir de tous les degrés d'appartenance le long de la branche concernée :

$$\alpha_N(\mathbf{x}_i) = \begin{cases} 1 & \text{si } N \text{ est la racine} \\ \alpha_M(\mathbf{x}_i) \wedge \mu_{A_j}(x_{ij}) & \text{sinon} \end{cases} \quad (1)$$

Si N est la racine, alors $\alpha_N(\mathbf{x}_i) = 1$ pour tout i .

Si N est un nœud ou une feuille, alors $\alpha_N(\mathbf{x}_i)$ est le produit de tous les degrés d'appartenance sur la branche, de la racine à N

avec les notations de la figure 2, où M est le parent de N , x_{ij} la variable attachée à M et N et A est un sous-ensemble flou sur le domaine de x_{ij} .

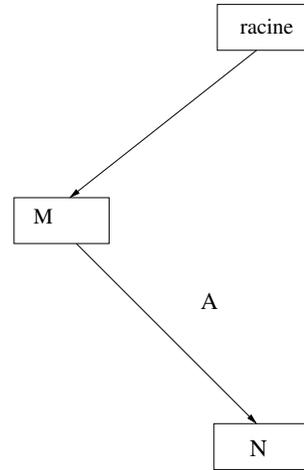


FIG. 2 – Appartenance au nœud N

– La sortie inférée par un ADDF est :

$$ADDF(\mathbf{x}) = \frac{\sum_F \alpha_F(\mathbf{x}) c_F}{\sum_F \alpha_F(\mathbf{x})} \quad (2)$$

– La conclusion associée à une feuille, c_F , est calculé par :

$$c_F = \frac{\sum_{j \in J} \alpha_F(\mathbf{x}_j) y_j}{\sum_{j \in J} \alpha_F(\mathbf{x}_j)} \quad (3)$$

¹Les arbres de régression englobent les arbres de décision. Leur sortie peut être un nombre réel quelconque et pas seulement un numéro de classe.

J : ensemble des couples (\mathbf{x}_j, y_j) pour lequel la valeur de vérité de la feuille est supérieure à un seuil donné.

3.2 Induction d'un ADDF

On parle d'induction car la construction permet de passer du particulier (les données de l'ensemble d'apprentissage) au général (les règles extraites des données). Cette construction procède par des partitions successives de l'ensemble d'apprentissage en sous-ensembles d'homogénéité croissante. Considérons le problème général suivant. Soit $f : (x_1, \dots, x_n) \rightarrow y = f(x_1, \dots, x_n)$ une relation d'entrée/sortie inconnue, les entrées x_1 à x_n étant des variables explicatives potentielles et y la sortie. Parmi les entrées, certaines sont importantes, d'autres redondantes, d'autres enfin inutiles. L'induction d'un ADDF va consister à :

- hiérarchiser les variables d'entrée en fonction de leur importance,
- évaluer l'utilité de prendre en compte ou non certaines entrées, soit en ligne, soit hors-ligne par élagage.

Il faut, pour cela, un ensemble d'apprentissage,

$E = \{(\mathbf{x}_i, \mathbf{y}_i); x_{ij} = (x_{i1}, \dots, x_{iM}), y_i \in \mathcal{R}, \text{ pour } i = 1 \text{ à } P\}$, où les x_{ij} , pour $j = 1$ à M ,

sont des variables linguistiques possédant m_j fonctions d'appartenance, $(A_{jk})_{j=1}^{m_k}$. Les fonctions d'appartenance forment des partitions floues fortes triangulaires sur les domaines d'entrée. Sans perte de généralité, la T-norme utilisée est le produit : $ET(x, y) = x \times y$. Ces choix apportent la propriété suivante :

$$\sum_F \alpha_F(\mathbf{x}) \equiv 1 \quad (4)$$

et l'équation 2 devient :

$$ADDF(\mathbf{x}) = \sum_F \alpha_F(\mathbf{x}) c_F \quad (5)$$

La construction automatique d'un arbre nécessite des mesures comme l'entropie et le gain d'information.

Soit $(B_k)_{k=1}^K$ un ensemble de fonctions d'appartenance sur le domaine de sortie. Sans perte de généralité, on peut supposer que ces fonctions d'appartenance forment une partition floue forte. Pour souligner l'analogie avec arbre de décision non flou, on peut considérer B_k comme une classe, pour $k = 1$ à K . On pose :

$\hat{\mu}_k(\mathbf{x}_i) = \mu_{B_k}(y_i)$: degré d'appartenance de \mathbf{x}_i à la classe k .

Dans un problème de classification, dans le cas idéal, les vecteurs d'apprentissage associés à un nœud terminal (une feuille) appartiennent à la même classe. On dit alors que la feuille est "pure". Ce n'est évidemment pas toujours possible et l'objectif de l'induction vise à créer des feuilles avec un degré de mélange minimum. Le partitionnement est réalisé à chaque nœud par des tests portant sur une variable : il faut donc choisir le meilleur test, en fonction d'une mesure d'efficacité dans la classification, le gain d'information.

- La notion de représentation de la classe k au nœud N pour une modalité A_j de la variable traitée, joue un rôle central. Elle est définie par :

$$r(k, j, N) = \sum_{i=1}^P \hat{\mu}_k(\mathbf{x}_i) \wedge \mu_{A_j}(\mathbf{x}_{ij}) \wedge \alpha_N(\mathbf{x}_i) \quad (6)$$

On en déduit les paramètres p_k et w_j utilisés pour le calcul du gain :

$$p_k = \frac{\sum_j r(k, j, N)}{\sum_k \sum_j r(k, j, N)} = \frac{\sum_j r(k, j, N)}{\sum_{i=1}^P \alpha_N(\mathbf{x}_i)} \quad (7)$$

$$w_j = \frac{\sum_k r(k, j, N)}{\sum_k \sum_j r(k, j, N)} = \frac{\sum_k r(k, j, N)}{\sum_{i=1}^P \alpha_N(\mathbf{x}_i)} \quad (8)$$

- Le gain d'information apporté par un attribut X au nœud N est :

$$G(X, N) = I(N) - \text{Info}(X, N) \quad (9)$$

avec

$$\begin{cases} I(N) = - \sum_k p_k \log p_k \\ \text{Info}(X, N) = \sum_j w_j I(X_j) \end{cases}$$

$I(N)$: Information au nœud N (entropie)

$\text{Info}(X, N)$: information apportée par X au nœud N

4 Implémentation du modèle

4.1 Variables explicatives

Le vecteur d'entrée du modèle ADDF est composé de trois descripteurs. A savoir :

- Intervalle RR_p : entrée E_0
- Rapport des intervalles $\frac{RR_s}{RR_p}$: entrée E_1
- Largeur du complexe QRS : entrée E_2

Le gain d'information de chaque variable à la racine de l'arbre, avant et après l'optimisation, est présenté dans le tableau 2.

TAB. 2 – Le gain d'information à la racine de l'arbre de chaque attribut avant et après l'optimisation.

Attributs	Avant-opt	Après-opt
E_0	0.16	0.33
E_1	0.11	0.34
E_2	0.34	0.62

4.2 Outil utilisé

Nous avons construit un arbre de décision flou (ADDF) à l'aide de la boîte à outil Wombat, avec scilab comme plateforme de calcul.

Wombat peut être utilisé en mode console ou avec une interface graphique écrite en Tcl/Tk. Il peut exporter l'arbre construit, soit sous une forme graphique (postscript, svg...), soit pour être utilisé indépendamment par un programme en langage C.

Les fonctions "bas niveau" de la boîte à outils sont calquées sur les fonctions théoriques introduites

dans (Glorennec 2003), Wombat permet ainsi aux utilisateurs de suivre pas-à-pas la croissance de l'arbre.

L'interface graphique, surcouchée des fonctions Scilab de plus haut niveau, permet à l'utilisateur de paramétrer la construction de son arbre en plusieurs étapes :

- configuration des variables grâce à un assistant ;
- modification des options et des valeurs modales de chaque variable, soit de façon numérique, soit de façon analogique ;
- visualisation des données d'apprentissage et de test.

Après cette phase de configuration des variables, l'algorithme d'optimisation permet à la fois de guider l'utilisateur dans le choix du nombre de sous-ensembles flous, et d'optimiser le placement des fonctions d'appartenance correspondantes.

La construction de l'arbre, enfin, est incrémentale : l'utilisateur peut décider de continuer ou d'arrêter le développement de l'arbre, en fonction d'indicateurs comme :

- le degré de mélange,
- l'erreur quadratique sur l'ensemble de validation, reflétant le pouvoir de généralisation de l'arbre construit,
- le niveau de complexité atteint.

pour obtenir un ADDF qui soit compact et lisible, il faut avoir le moins possible de fonctions d'appartenance, ce qui risque de nuire aux performances. Il faut donc optimiser.

Le principe consiste à ajuster les fonctions d'appartenance sur une entrée, pour optimiser le gain d'information, dans deux cas :

- avec le nombre de fonctions d'appartenance fixé par l'utilisateur,
- avec une fonction d'appartenance de plus.

L'utilisateur choisit la structure optimale en prenant en compte l'accroissement relatif du gain d'information.

5 Analyse et discussion des résultats

Les résultats du tableau 3 montre que l'arbre produit globalement le bon diagnostic avec un certain degré d'erreur. Une analyse plus approfondie montre que, dans tous les cas, on peut trouver des cycles différents des BVPs et présentant exactement les mêmes caractéristiques, par rapport aux 3 variables considérées. Chaque feuille signifie que l'on peut être capable d'identifier la classe la plus représentative dans une base d'exemples. Lorsque cette feuille est pure, il n'y a qu'une seule classe possible et c'est elle qui étiquette la feuille. C'est le cas idéal mais il est rarement satisfait en pratique. Dans tous les autres cas, la règle générale consiste à étiqueter la feuille avec la classe possédant le plus de représentants. Dans notre expérimentation, chaque feuille est caractérisé par :

- des degrés d'appartenance maximum à la feuille pour un élément,
- un nombre pondéré d'éléments, ie chaque vecteur de la base d'apprentissage peut être représenté avec un certain pourcentage dans plusieurs feuilles de l'arbre.

Ces paramètres constituent simplement une source d'information sur la qualité de la discrimination au niveau de chaque feuille :

- La feuille 0 représente environ 60 battements 'Nm' avec une activation supérieure à 90 % par contre un seul battement 'Bvp' est représenté avec une activation supérieure à 60 %. Elle est quasiment pure,
- La feuille 1 représente environ 200 battements 'Nm' avec une activation supérieure à 90 % par contre aucune représentation des battements 'Bvp' (activation inférieure à 20 %). Elle est pure,
- La feuille 2 représente un mélange dont le 1/3 est de type 'Nm' et le 2/3 est de type 'Bvp' : il manque une variable explicative, car certains battements 'Nm' (peu de cas) sont indiscernables des 'BVPs',
- La feuille 3 ne représente que des battements de type 'Nm', cette feuille est donc pure,
- La feuille 4 représente un mélange mais majoritairement des 'BVPs',
- La feuille 5 représente environ 200 battements 'Bvp' avec une activation supérieure à 90 % par contre 2 battements 'Nm' sont représentés avec une activation supérieure à 70 %. Elle est quasiment pure,
- La feuille 6 représente un mélange où chaque type est représenté à 50 %, on ne peut rien dire : heureusement ceci ne porte que sur peu de cas,
- La feuille 7 représente seulement deux battements Bvp avec une activation supérieure à 60 %, cette feuille est presque pure.

TAB. 3 – Nombre pondéré d'éléments (np) et degrés d'appartenance maximum à la feuille ($dapp$) pour un élément

N^0 de la feuille	classe N		classe BVP	
	np	$dapp$	np	$dapp$
feuille 0	90.79	1	2.33	0,68
feuille 1	255.25	1	0.42	0,19
feuille 2	8.24	1	15.33	1
feuille 3	13.31	1	0.18	0,11
feuille 4	10.65	1	88.85	1
feuille 5	7.81	1	253.40	1
feuille 6	20.77	1	17.27	1
feuille 7	2.14	0,67	12.17	0,91

Nous remarquons une ambiguïté pour les feuilles 2 et 6 mais peu de battements sont concernés, nous pouvons donc simplifier l'arbre en gardant uniquement les feuilles représentatives, la figure 3 présente le modèle de l'arbre simplifié avec six feuilles seulement. Les erreurs quadratiques moyennes sur l'ensemble d'apprentissage et l'ensemble de test étaient : $EQM(app) = 0.044$ et $EQM(test) = 0.034$.

6 Conclusion

Le travail présenté dans cet article est essentiellement axé sur l'élaboration d'un classifieur flou en utilisant

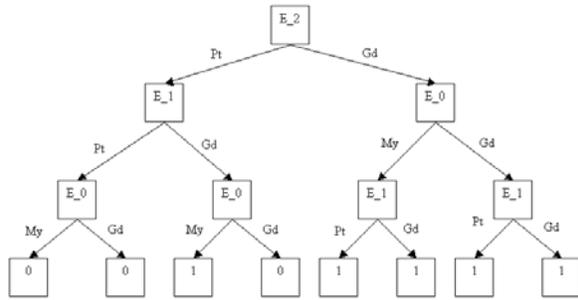


FIG. 3 – Arbre de décision flou simplifié pour la reconnaissance des BVP (les feuilles sont numérotées de gauche à droite)

les arbres de décision flous (ADDFs). Cette approche permet de pallier au problème d'interprétabilité des résultats rencontré dans plusieurs approches (réseaux de neurones, réseaux bayésiens etc ...). La contribution des modèles ADDF a porté essentiellement sur la hiérarchisation des variables d'entrée (indiquer les variables les plus importantes) et leur acceptation d'un nombre plus élevé d'entrées par rapport aux autres modèles (systèmes d'inférence floue etc ...). Dans cette approche, nous avons montré l'intérêt et l'impact des modèles à base des règles sur les techniques de classification automatique utilisées dans le domaine médical.

Il n'en reste cependant pas moins que ce travail constitue les fondements d'un travail à poursuivre. Les perspectives futures sont en premier lieu la recherche et l'extraction d'autres descripteurs et indicateurs pertinents du signal électrique du cœur, par une approche utilisant des techniques de transformations du signal et de reconnaissance de formes. Sachant qu'une bonne caractérisation de l'ECG conditionne la réussite de la phase de classification.

Références

[Antonia 2003] Antonia V., and al., "Diagnosis of Ovarian Cancer Using Decision Tree Classification of Mass Spectral Data", *Journal of Biomedicine and Biotechnology*, pp. 105-109, 2003.

[Carrault 2003] Carrault, G., Cordier, M.O., Quiniou, R., and Wang, F., "Temporal abstraction and Inductive Logic Programming for arrhythmia", *Artificial Intelligence in Medicine, Elsevier*, vol. 28, p. 231-263, 2003.

[Chidanand 1997] Chidanand A., and Sholom W., "Data mining with decision trees and decision rules", *Future Generation Computer Systems* ., 13(2-3), pp.197-210, 1997.

[Chow 1995] Chow, H., Chang, W., "QRS feature extraction", *Proc. of Computers in Cardiology, IEEE Computer Society*, pp. 659-662, 1992.

[Giesen 1996] Giesen R.J, and al., "Construction and application of hierarchical decision tree for classification of ultrasonographic prostate images", *Med. Biol. Eng. Comput*, pp. 105-109, 1996.

[Glorennec 2003] Glorennec P.Y., and al., "Constrained optimisation of fuzzy decision trees, in Interpretability Issues in Fuzzy Modeling",

[Lagerholm 2000] Lagerholm, M., and al., "Clustering ECG complexes using hermite functions and self-organizing maps", *IEEE Trans. Biomed. Eng.* , pp. 838-848, 2000.

[Le Guillou 2004] Le Guillou, X., Glorennec, P.Y., "Wombat : a toolbox to build and optimize a fuzzy regression tree ", *First. Int. Conf. SCILAB04, Rocquen-court*, Dec 2004.

[Mia 2003] Mia K., and al., "Decision Tree classification of proteins identified by mass spectrometry of blood serum samples from people with and without lung cancer", *Proteomics 2003*, pp 1678-1679, 2003.

[Rasoul 1991] Rasoul S., and David L., "A survey of decision tree classifier methodology", *IEEE Transactions on Systems, Man, and Cybernetics* , 21(3), pp.660-674, 1991.

[Shu 2004] Shu T., and Chao-Tung Y., "Decision Tree Construction for Data Mining on Grid Computing", *IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04)*, March 28-31, Taipei-Taiwan, 2004.

[Silipo 1999a] Silipo R., and al., "Design of hybrid architectures based on neural classifier and RBF pre-processing for ECG analysis", *International Journal of Approximative Reasoning*, 21, pp.177-196, 1999.

[Sotiris 2004] Sotiris A.P, and al., "Decision Tree-based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds", *BioMedical Engineering Online* , pp. 3-21, June 2004.

[Sreerama 1998] Sreerama K.M., "Automatic construction of decision trees from data : a multi-disciplinary survey", *Data Mining and Knowledge Discovery*, 2(4), pp.345-389, 1998.

[Suzuki 1995] Suzuki Y., "Self-Organizing QRS-Wave Recognition in ECG Using Neural Networks", *IEEE Transactions on Neural Networks*, Vol. 6, Issue 6, pp. 1469-1477, 1995.