

Analyse de données pour la construction de modèles de procédures neurochirurgicales

Brivael Trelhu, Florent Lalys, Laurent Riffaud, Xavier Morandi, Pierre Jannin

▶ **To cite this version:**

Brivael Trelhu, Florent Lalys, Laurent Riffaud, Xavier Morandi, Pierre Jannin. Analyse de données pour la construction de modèles de procédures neurochirurgicales. Extraction et gestion des connaissances (EGC 2009), Jan 2009, Strasbourg, France. pp.427-432. inserm-00616975

HAL Id: inserm-00616975

https://inserm.hal.science/inserm-00616975

Submitted on 25 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse de données pour la construction de modèles de procédures neurochirurgicales

Brivael Trelhu[a], Florent Lalys[a], Laurent Riffaud[a,b], Xavier Morandi[a,b], Pierre Jannin[a]

[a] IRISA, U746 VisAGeS, 2, Avenue du Pr. Léon Bernard CS 35043, Rennes Cedex, France
brivael.trelhu@irisa.fr, florent.lalys@irisa.fr, pierre.jannin@irisa.fr
[b] Department of Neurosurgery, Hopital Universitaire de Rennes, 2, Avenue du Pr. Léon Bernard CS 35043, Rennes Cedex, France
laurent.riffaud@chu-rennes.fr, xavier.morandi@chu-rennes.fr

Résumé. Dans cet article, nous appliquons une méthode d'analyse sur des descriptions de procédures de neurochirurgie dans le but d'en améliorer la compréhension. La base de données XML utilisée dans cette étude est constituée de la description de 157 chirurgies de tumeurs. Trois cent vingt deux variables ont été identifiées et décomposées en variables prédictives (connues avant l'opération) et variables à prédire (décrivant des gestes chirurgicaux). Une analyse factorielle des correspondances (AFC) a été réalisée sur les variables prédictives, ainsi qu'un arbre de décision basé sur un dendrogramme préalablement établi. Six classes principales de variables prédictives ont ainsi été identifiées. Puis, pour chacune de ces classes, une analyse AFC a été réalisée sur les variables à prédire, ainsi qu'un arbre de décision. Bien que le nombre de cas et le choix des variables constituent une limite à cette étude, nous avons réussi à prédire certaines caractéristiques liées aux procédures en partant de données prédictives.

1 INTRODUCTION

La compréhension des processus décisionnels lors de la réalisation d'une procédure chirurgicale nécessite de s'appuyer sur une description explicite de celle-ci. De nombreux ouvrages chirurgicaux et revues ont décrit les principales procédures (voies d'abord, techniques d'exérèse…), complétés par une multitude d'articles de notes techniques (Rhoton 2003, Hernesniemi et al. 2005) à partir d'avis d'experts ou de cas cliniques. Ces descriptions sont réalisées par étude manuelle de cas. Il n'existe pas (ou très peu) de travaux permettant la création de tels modèles génériques des procédures chirurgicales à partir de modèles *patient-spécifiques* décrivant des cas chirurgicaux. Pour cela, il est nécessaire de disposer de descriptions formelles et explicites des procédures réalisées comprenant à la fois les informations disponibles avant l'opération (données prédictives) mais aussi les différentes techniques utilisées complétées par l'utilisation des instruments (en tenant compte de leur rôle et de leur impact sur le déroulé chirurgical).

La procédure de chirurgie endoscopique a été décomposée en étapes, sous étapes et tâches hiérarchiques successives et modélisée sous forme graphique (Cao et al. 1999). Cette représentation a été suggérée pour mesurer les performances chirurgicales entre chirurgiens

expérimentés et novices, pour évaluer et comparer deux techniques chirurgicales différentes pour le même type d'intervention et pour améliorer les performances du chirurgien par l'entraînement sans toutefois détailler chacune de ces propositions. Le modèle prévu de planification de la procédure chirurgicale par le chirurgien se faisait à partir d'un modèle défini de procédure-type tiré d'un catalogue à laquelle était rattachée une base de données d'informations techniques et de connaissances médicales facilement accessible. La même démarche avec accès aux informations était imaginée pendant la procédure elle-même. Les buts étaient multiples et évidents : aider le chirurgien dans ses décisions de planning préopératoire et pendant la phase d'exécution, clarifier le déroulé de la procédure chirurgicale (Münchenberg et al. 2000, Tonetti et al. 2007). Cependant, dans la très grande majorité des travaux réalisés, les auteurs se sont focalisés sur des procédures chirurgicales à faible variabilité inter patient (comme la laparoscopie).

Dans le but de mieux connaître et décrire une procédure chirurgicale et donc d'améliorer sa préparation et sa réalisation, une approche consiste à développer des outils permettant 1) de décrire de façon structurée des cas chirurgicaux, 2) de stocker ces descriptions dans des bases de données et 3) d'extraire les connaissances (ou modèles génériques) à partir de l'analyse de ces bases de données. Des équipes se sont focalisées sur les deux premières étapes pour le développement d'outils de description de procédures chirurgicales en amont et en aval de la procédure (Jannin et al. 2003,2007, Tonetti et al. 2007) mais aussi pendant celle-ci (Neumuth et al. 2006).

Dans cet article, nous proposons une approche pour tendre vers la génération des modèles génériques de procédures chirurgicales. Le but de l'étude présentée est de mettre en évidence des paramètres importants permettant de discriminer différentes procédures en neurochirurgie tumorale. Ainsi nous avons déterminé l'importance de certains critères lors de la réalisation des procédures, permettant ainsi la création d'arbres de décision.

2 MATERIELS ET METHODES

Un travail précédent ayant permis une première modélisation des procédures de neurochirurgie, nous avons ensuite postulé que ces procédures étaient fractionnables en différentes étapes distinctes (Jannin et al. 2007).

Notre étude a été réalisée sur une base de 157 procédures d'exérèse de tumeurs des hémisphères cérébraux choisies du fait de leur fréquence de réalisation dans le département de neurochirurgie du CHU de Rennes. Les enregistrements de ces procédures ont été réalisés rétrospectivement par les neurochirurgiens eux-mêmes selon un protocole précédemment établi (Jannin et al. 2003). Il s'agit d'enregistrements décrivant les différentes étapes des procédures, complétées par des paramètres propres à chaque opération : informations relatives au patient (comme l'âge, le sexe, et la latéralité) mais aussi particularités de chaque cas (comme la localisation de la tumeur, sa bénignité ou malignité). Ces enregistrements ont été réalisés à l'aide d'un logiciel développé au sein de l'équipe, basé sur une interface web développée avec le langage PHP. Les descriptions des procédures ont été stockées dans des fichiers XML. La base de données des 157 enregistrements dont nous disposions nous a permis d'extraire 322 variables sur lesquelles a reposé notre étude. A partir de ces 322 variables, nous avons construit un tableau binaire décrivant la présence de chaque variable pour chaque enregistrement. L'étape suivante a été la discrimination des variables prédictives et des variables à prédire, définies comme suit : 1) toute variable connue avant l'opération

était une variable prédictive, de même, 2) toute variable faisant partie intégrante de la procédure neurochirurgicale était une variable à prédire. Les variables prédictives concernaient la description du cas chirurgical (âge et sexe du patient, localisation et type de la tumeur, …). Les 322 variables se subdivisaient en 152 variables prédictives et 170 variables à prédire. Toutes les variables extraites étaient binaires : présence ou absence de la variable pour chaque enregistrement de procédure. Le traitement et l'analyse des données du tableau binaire ont été réalisés sous l'environnement Matlab. De nombreuses solutions ont été testées afin d'obtenir des résultats satisfaisants. Nous avons tout d'abord réalisé une analyse factorielle des correspondances (AFC) basée sur les variables prédictives. Une première sélection de variables, obligatoire pour le calcul de l'AFC, a ainsi été effectuée. L'analyse factorielle est calculée par l'estimation du maximum de vraisemblance (MLE) de la matrice des facteurs dans le modèle d'analyse factoriel :

$$x = \mu + \Lambda f + e \qquad (1)$$

où x est le vecteur des variables observées, μ est le vecteur constant des moyennes, Λ est la matrice des covariances entre les variables observables et les facteurs communs, f est le vecteur des facteurs communs, e est le vecteur des facteurs spécifiques indépendants.
L'étape suivante est le calcul de la distance euclidienne entre les paires d'objets dans la matrice des données x, définit comme suit :

$$d_{rs}^2 = (x_r - x_s)\,(x_r - x_s)' \qquad \text{(avec } x_r \text{ et } x_s \text{ 2 objets de } x\text{)} \qquad (2)$$

Puis une opération de couplage (linkage) a été réalisée en fonction des distances euclidiennes. Elle crée un arbre de clusters hiérarchiques des distances à partir de la matrice des distances euclidiennes obtenues précédemment. La méthode employée est celle du plus proche voisin (ou single linkage), qui détermine la plus petite distance entre les objets de deux clusters :

$$d(r,s) = min\,(dist(x_{ri}, x_{sj})),\ i \in (1, \ldots, n_r),\ j \in (1, \ldots, n_s) \qquad (3)$$

où r est le cluster formé, n_r le nombre d'objets dans le cluster r, x_{ri} est le $i$ème objet dans le cluster r.
Puis un dendrogramme est créé à partir des résultats du couplage. Pour déterminer le nombre de clusters optimal pour l'arbre de décision, nous avons choisi une valeur seuil de significativité, à savoir 70% de valeur maximale du couplage obtenue lors de l'étape précédente, à partir de laquelle nous avons généré les clusters significatifs.
L'arbre de décision a alors été généré en fonction de ces clusters.
Une fois cette étape sur les variables prédictives effectuées, et pour chaque cluster correspondant aux extrémités de l'arbre, nous avons déterminé la population de ces clusters puis nous avons réalisé une étude identique sur les variables à prédire. Ainsi nous avons obtenu un arbre de décision construit à partir des profils de variables à prédire des enregistrements composant le cluster.

Nous avons ensuite cherché à déterminer s'il y avait une corrélation entre les arbres prédictifs et les arbres à prédire. Pour se faire, nous avons extrait des arbres de décision propres à chaque cluster, les profils des variables prédictives et des variables à prédire. Nous avons demandé ensuite à un expert neurochirurgien son avis sur les relations entre les variables prédictives et les variables à prédire qui ont été mises en valeur par l'étude.

3 RESULTATS

Les 322 variables extraites à partir des enregistrements XML des procédures n'étaient pas toutes exploitables, et ce de part l'AFC elle-même. En effet, il n'a pas été pas possible de réaliser une AFC si la présence ou non d'une variable selon les différents enregistrements était une combinaison linéaire de la présence d'une autre variable. Ainsi 58% des variables se sont avérées inutilisables. Au final, 31 variables prédictives et 107 variables à prédire ont été utilisées.

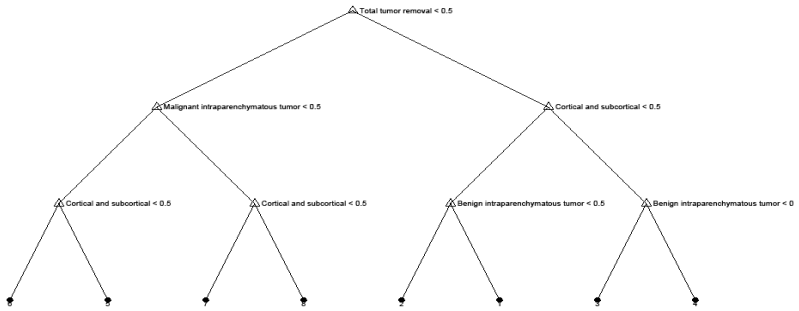


FIG. 1 – *Arbre de décision basé sur les variables prédictives des 157 cas.*

Le critère majeur qui a permis de faire la distinction entre les différents clusters était le type d'opération. Le type de tumeurs ainsi que l'emplacement de la tumeur ont également un rôle prépondérant.

Cluster n°	nb de procédures	Variables prédictives			
		Total tumor removal	Cortical and Subcortical	Malignant tumor	Benign tumor
1	19	oui	non	/	oui
2	60	oui	non	/	non
3	34	oui	oui	/	non
4	15	oui	oui	/	oui
5	3	non	oui	non	/
6	15	non	non	non	/
7	18	non	non	oui	/
8	12	non	oui	oui	/

TAB.1 – *Caractéristiques communes de chaque cluster de variables prédictives.*

En ce qui concerne le cluster n°5 [Total tumor removal -, Cortical and Subcortical +, Malignant tumor -], nous ne disposions que de 3 enregistrements, ce qui rendait l'étude irréalisable.

Cluster n°	nb de procédures	variables à prédire				
		discrimination principale	branche gauche		branche droite	
			1e nœud	2e nœud	1e nœud	2e nœud
1	19	superior temporal sulcus	area to be avoided	/	/	/
2	60	U shape (skin incision)	/	/	Horseshoes shape (duramater incision)	/
3	34	H shape (skin incision)	superior frontal sulcus	anatomic and fonctional area identification	superior frontal sulcus	anatomic and fonctional area identification
4	15	anatomic and fonctional area identification	/	/	ultrasound aspiration	arciform
5	3	/	/	/	/	/
6	15	craniotome	/	/	/	/
7	18	H shape (skin incision)	/	/	anatomic and fonctional area identification	/
8	12	H shape (skin incision)	/	/		/

TAB.2 – *Résultats obtenus à partir des clusters de variables prédictives.*

4 DISCUSSION

Dans cette étude, nous avons appliqué une méthode d'analyse sur 157 procédures d'exérèse de tumeurs des hémisphères cérébraux. Nous avons pu identifier des groupes de procédures en fonction de variables décrivant le cas chirurgical et connues avant l'intervention. Pour chaque groupe, nous avons identifié des variables discriminantes qui sont des caractéristiques de la procédure chirurgicale.

Le cinquième cluster ne contenait que trois cas chirurgicaux. De ce fait, bien que l'on puisse corréler les profils obtenus à partir des variables prédictives avec les variables à prédire les concernant, cela n'avait pas de signification statistique. Quatre clusters avaient le type d'incision comme discrimination principale. Ainsi l'importance du type d'incision dans le déroulement d'une procédure chirurgicale a été déductible de ces résultats. Les sillons corticaux apparaissaient également dans plusieurs arbres. L'importance des sillons dans la réalisation de ce type de procédure neurochirurgicale a été ainsi mise en avant. En tenant compte du profil de chaque cluster pour les variables prédictives, on a cherché à vérifier et à déterminer s'il y avait une corrélation entre ces dernières et les variables à prédire. Les résultats de cette étude ont été présentés à un expert neurochirurgien. Pour chaque cluster, il a vérifié la cohérence des règles que nous avons obtenues. Par exemple, le 1^{e} cluster caractérisait des tumeurs bénignes qui étaient dans la zone temporale supérieure (1^{e} classe), proches du sillon temporal supérieur (2^{e} classe) et dans la région centrale du cerveau, en zones éloquentes (3^{e} classe). L'expert neurochirurgien a ainsi justifié toutes les classifications obtenues. Le 3^{e} cluster apportait une validation supplémentaire : il y avait une cohérence entre les deux premières classes de ce cluster. Les zones de localisation dans le cerveau qu'elles caractérisaient étaient proches, ainsi la classification obtenue était cohérente.

L'étude des variables prédictives a permis de mettre en évidence différentes façons de procéder pour une même situation. Elle a permis de définir des grands cadres de pathologies : clusters de tumeurs malignes, clusters de tumeurs bénignes qui dégénèrent, ou encore clusters de tumeurs bénignes. Elle a aussi fait apparaître des clusters de qualité d'exérèse. Au niveau des variables à prédire, le sillon temporal supérieur a été une variable plusieurs fois présente, son importance dans la neurochirurgie étant ainsi mise en avant. L'étude a aussi démontré une variation de l'utilisation de la phase d'identification des zones anatomiques et fonctionnelles. En effet, si l'utilisation de cette dernière était hétérogène dans les cas de tumeurs bénignes qui dégénèrent, ce n'était pas le cas pour les autres types de tumeurs puisque la variable n'est pas retenue. Sa répartition a été donc homogène dans ces populations. Ainsi soit cette phase d'identification a toujours été présente, soit elle ne l'a jamais été. Selon le neurochirurgien senior, l'utilisation du repérage des zones anatomiques et fonctionnelles dans les cas de tumeur maligne n'est pas habituelle, par exemple en raison du caractère d'urgence de la chirurgie. A l'opposé, dans les cas de tumeurs bénignes, ces zones sont presque toujours identifiées. De ce fait, les temps de préparation de la procédure chirurgicale sont plus longs dans les cas de tumeurs bénignes car ils nécessitent davantage d'examens (IRM 3D, IRMf,…), d'analyses et de repérage.

Les résultats de cette étude comportent des limites. Tout d'abord, une analyse factorielle des correspondances a été réalisée sur les variables prédictives. Pour cela, un tri a été nécessaire au calcul de l'AFC. En effet si deux variables ont le même profil d'appartenance, c'est-à-dire que les deux lignes du tableau étaient identiques, ou complémentaires l'une de l'autre, l'AFC ne pouvait pas être calculé. Ce tri a orienté l'analyse, et bien qu'obligatoire, la sélection d'une variable entre deux a constitué une limite à l'étude.

Bien que le nombre de procédures et le choix des variables aient constitué une limite à cette étude, nous avons pu identifier des variables prédictives de certains gestes chirurgicaux. Cette étude n'est qu'une première étape. L'approche devra être plus robuste et étendue à la globalité des caractéristiques de la procédure chirurgicale. Il sera aussi nécessaire de travailler sur une représentation intuitive des résultats qui puisse directement être utilisée par les neurochirurgiens pour l'anticipation des interventions.

5 REFERENCES

Cao CG., MacKenzie CL., Ibbotson JA., Turner LJ., Blair NP., Nagy AG.(1999). Hierarchical decomposition of laparoscopic procedures. Stud Health Technol Inform., 62:83-9.

Hernesniemi J, Niemelä M, Karatas A et al. (2005). Some collected principles of microneurosurgery: simple and fast, while preserving normal anatomy. A review. Surg Neurol, 64:195-200.

Jannin P, Raimbault M, Morandi X, Riffaud L, Gibaud B. (2003). Models of surgical procedures for multimodal image-guided neurosurgery. Comput Aided Surg, 8:98-106.

Jannin P, Morandi X (2007). Surgical models for computer-assisted neurosurgery. Neuroimage, 37:783-791.

Münchenberg J, Brief J, Raczkowsky J, Wörn H, Hassfeld S, Mühling J (2000). Operation planning of robot supported interventions. In: International Conference on Intelligent Robots and Systems. Takamatsu, Japan: IEEE/RSJ, Inc., p547-52.

Neumuth T, Durstewitz N, Fischer M, et al. (2006). Structured recording of intraoperative surgical workflows. In Horii Sc, ratib OM. SPIE Medical Imaging : PACS and Imaging Informatics 7 (31); SPIE Bellingham 2006; CID 61450A

Rhoton AL (2003). Operative techniques and instrumentation for neurosurgery. Neurosurgery, 53:907-34.

Tonetti J., Vadcard L.,Luengo V., Dubois M. (2007). Contributions interdisciplinaires à la conception d'un environnement de formation en orthopédie. Surgetica

Summary

In this article, we present a study for the analysis of neurosurgical procedures with the aim of improving their comprehension. XML Database used in this study included 157 tumors surgeries descriptions. 322 variables were identified and broken down into predictive variables (i.e., known before operation) and into variables to be predicted (i.e., pertaining to the surgical gestures). A correspondence factorial analysis (CFA) was carried out on the predictive variables, as well as a build of a decision tree based on a beforehand dendrogram established. Six principal classes of predictive variables were thus identified. For each one, CFA was carried out on the variables to be predicted, as well as a decision tree. Although the number of limited cases and the variables choice constitute a limit to this study, we succeeded in predicting variables related to surgery on the basis of predictive data.