

Classification pour l'aide au traitement du cancer du sein

Benoît Bresson, Jean Lieber

► **To cite this version:**

Benoît Bresson, Jean Lieber. Classification pour l'aide au traitement du cancer du sein. Florence Lieber, Jean-Francois Mari, Amedeo Napoli, Arnaud Simon. Septième journées de la Société Francophone de Classification - SFC'99, 1999, Nancy, France, Unité de recherche INRIA Lorraine, 7 p, 1999. <inria-00098763>

HAL Id: inria-00098763

<https://hal.inria.fr/inria-00098763>

Submitted on 26 Sep 2006

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.

Classification pour l'aide au traitement du cancer du sein

Benoît Bresson et Jean Lieber
LORIA, UMR 7503 — CNRS, INRIA-Lorraine, Universités de Nancy
BP 239,
54506 Vandœuvre-lès-Nancy,
mél. : {bresson,lieber}@loria.fr

Résumé

Cet article présente le système CASIMIR, un système à base de connaissances en cours de conception et de développement qui a pour objectif l'aide à la décision thérapeutique dans le cadre du traitement du cancer du sein. Les connaissances sont représentées par une hiérarchie de classes définies par les propriétés utiles des patients (et de leurs tumeurs). À certaines de ces classes sont associés des traitements médicaux. Le raisonnement dit protocolaire consiste à classer le patient à traiter et les propriétés de son cancer dans cette hiérarchie puis à en déduire les traitements à effectuer. Malheureusement, un tel raisonnement ne donne pas nécessairement satisfaction. En effet, d'une part les classes dont on dispose sont insuffisantes pour traiter tous les cas et, d'autre part, même si le raisonnement protocolaire fournit un traitement, ce traitement peut s'avérer inapplicable. Par conséquent, un raisonnement non protocolaire s'avère nécessaire. Ce raisonnement consiste à adapter les classes considérées et les traitements associés au patient. Les connaissances mises en jeu pour l'adaptation sont destinées à être intégrées au protocole, avec l'accord des spécialistes.

Mots-clés classification dure, classification élastique, raisonnement à partir de cas, aide à la décision, cancer du sein. .

1 Introduction

Le traitement du cancer du sein est un domaine complexe qui fait intervenir de nombreux paramètres liés au patient à traiter, en particulier son âge, son sexe, ses antécédents, les caractéristiques de sa ou de ses tumeurs, etc. La première phase du raisonnement des oncologues est *protocolaire*, c'est-à-dire s'appuyant sur une méthode systématique : le protocole. Si cette phase échoue, un raisonnement non protocolaire lui succède.

Cet article présente CASIMIR, une application en cours de conception et de développement qui a pour objectif l'aide au traitement du cancer du sein. Le raisonnement protocolaire peut être considéré comme un raisonnement à base de règles. Nous le modélisons par une classification « dure » dans une hiérarchie qui constitue la base des connaissances protocolaires. Si, pour une raison ou une autre, ce raisonnement ne donne pas satisfaction, un raisonnement non protocolaire doit le remplacer. Celui-ci peut être considéré comme un raisonnement à partir de cas et s'appuie sur une classification « élastique » dans la hiérarchie.

Après avoir défini au paragraphe 2 quelques notions concernant le traitement du cancer du sein, nous expliquerons les principes de notre modélisation du raisonnement des oncologues : nous aborderons le raisonnement protocolaire au paragraphe 3 et le raisonnement non protocolaire au paragraphe 4. Enfin, nous expliquerons au paragraphe 5 comment cette modélisation peut aider à faire évoluer le protocole.

Au stade actuel, le raisonnement protocolaire est en cours d'implantation. En revanche, le raisonnement non protocolaire et les outils d'aide à l'évolution du protocole n'en sont qu'à une phase de réflexions préliminaires.

2 Le traitement du cancer du sein

Dans le cadre du traitement du cancer du sein, les médecins disposent d'un « protocole de traitement des carcinomes mammaires au stade locorégional ». Ce protocole est constitué d'un ensemble de règles qui ont été établies grâce à des études statistiques. Tout au long du traitement du cancer, le médecin prend des décisions sur la thérapeutique à suivre. Ces décisions lui sont dictées par ce protocole, tant au niveau des actes chirurgicaux qu'au niveau des traitements médicaux. Malheureusement, une partie des cas de cancers du sein (environ 30%), n'est pas prise en compte par ce protocole, soit parce que ces cancers sont atypiques et que le protocole ne permet pas de leur associer de traitement soit parce que le médecin a des scrupules à appliquer le traitement donné par le protocole. Pour pallier à ces problèmes, un groupe d'experts, le *comité de décision thérapeutique*, est consulté pour résoudre ces cas non protocolaires.

3 Raisonnement protocolaire et classification dure

Le raisonnement protocolaire est modélisé grâce à la classification dure dont le principe est décrit au paragraphe 3.1 et dont un exemple est présenté au paragraphe 3.2. Ce raisonnement s'appuie sur des connaissances détenues par les cancérologues et qu'il faut acquérir, comme nous le verrons au paragraphe 3.3. Le paragraphe 3.4 explique le lien de cette modélisation du raisonnement avec le système RÉSYN dédié à la chimie organique.

3.1 Principe de la classification dure

Le protocole est constitué d'un ensemble de règles $R = \text{Prém} \rightarrow \text{Ccl}^\circ$ où Prém est la prémisse de la règle R et Ccl° , sa conclusion. Plus précisément, Prém est un ensemble de conditions susceptibles d'être vérifiées par un patient et Ccl° est un traitement (ou un ensemble de traitements) associé à ces conditions.

Prém peut être représenté par une classe CL dont l'ensemble des propriétés — ou *intension* — est Prém et qui dénote l'ensemble de tous les patients possibles vérifiant ces conditions (cet ensemble est aussi appelé *extension* de la classe). À la classe CL on associe le traitement Ccl° . Ainsi, la connaissance de la classe CL équivaut à la connaissance de la règle R : Prém est l'intension de CL et Ccl° est le traitement associé à CL .

Le patient est lui aussi décrit par une classe, dénotée CL_{patient} , dont l'intension est l'ensemble des propriétés connues du patient qui sont éventuellement utiles à la décision (son sexe, son âge, les caractéristiques de sa ou de ses tumeurs, etc.). Notons que le patient est décrit par une classe plutôt que par une instance car toutes ses propriétés ne sont pas nécessairement connues : on peut ne pas connaître son âge, on peut ne connaître la taille de sa tumeur que de façon imprécise, etc. Préciser ces caractéristiques revient donc à spécialiser la classe CL_{patient} . Décrire le patient dans un formalisme de classe permet en outre de le comparer plus aisément avec les connaissances décrites par une hiérarchie de classes.

Tester si la règle $R = \text{Prém} \rightarrow \text{Ccl}^\circ$ peut être appliquée au patient revient à tester si les conditions de Prém sont satisfaites par le patient. Cela revient donc à tester si la classe CL_{patient} est *plus spécifique* que la classe CL associée à R . Rappelons qu'une classe CL_1 est plus spécifique que

qu'une classe CL_2 si l'extension de CL_1 est incluse dans l'extension de CL_2 ou, de façon équivalente, si la conjonction des propriétés de l'intension de CL_1 entraîne la conjonction des propriétés de l'intension de CL_2 . On note alors $CL_1 \preceq CL_2$.

Les n classes CL correspondant aux n règles du protocole sont organisées hiérarchiquement pour l'ordre \preceq . Le sommet de la hiérarchie est la classe ω dont l'extension est l'ensemble de tous les patients possibles et l'intension est la propriété « toujours vraie ». Effectuer la classification de CL_{patient} dans la hiérarchie permet de mettre en évidence les classes CL plus générales que CL_{patient} . Le traitement associé à ces classes va permettre de donner le traitement proposé par le protocole : ce sera la conjonction des traitements associés aux classes CL . En cas de conflit entre les traitements associés aux classes CL_1 et CL_2 plus générales que CL_{patient} , si $CL_1 \preceq CL_2$, alors c'est le traitement associé à CL_1 qui est choisi, selon le principe du masquage dans le cadre des systèmes de représentation des connaissances par objets [Masini *et al.*, 1989]. La hiérarchie est faite de telle sorte que deux classes incomparables pour \preceq ne peuvent entraîner de conflit. Nous appelons cette méthode de résolution de problèmes la *classification dure* (par opposition à la classification élastique, présentée au paragraphe 4).

3.2 Exemple

La figure 1 illustre un exemple de classification dure. Les classes CL_4 à CL_7 constituent une partie de la hiérarchie (qui contient en fait plusieurs centaines de classes). Chaque classe est décrite par des caractéristiques de la patiente, de sa tumeur et — sous la barre horizontale de chaque classe — par un traitement. Par exemple « PATEY = Oui » indique qu'un « Patey » doit être fait (traitement chirurgical non conservateur qui consiste en une ablation totale du sein). « PATEY = ? » indique qu'un « Patey » est envisagé et qu'il faut décider si oui ou non il doit être effectué. La classe CL_{patient} représentant la patiente à traiter « descend » le long de la hiérarchie (les flèches $t = 1$ à $t = 5$ symbolisent cette descente). Ce raisonnement conclut donc qu'un Patey doit être fait. D'autres traitements peuvent être également mis en évidence par la suite du raisonnement correspondant à la poursuite de la classification dure (dans la partie de la hiérarchie non représentée sur la figure) : plus CL_{patient} descend dans la hiérarchie et plus le traitement proposé est complet.

3.3 Acquisition des connaissances sous forme hiérarchique

Cette modélisation de la connaissance protocolaire ne va pas sans difficultés : faire parler un groupe d'experts est rarement facile. En l'occurrence, la démarche utilisée est la suivante.

Le protocole de traitement utilisé par les médecins est transcrit sur une plaquette dont le principal objectif est de décrire, d'une façon relativement succincte, les grandes lignes du protocole. Bien que cette plaquette ne soit pas exhaustive, elle sert de point de départ pour l'acquisition des connaissances.

Cette acquisition consiste dans un premier temps à établir le vocabulaire utile à la prise de décision protocolaire qui permettra de définir les propriétés des classes (âge, sexe, taille de la tumeur, etc.). Puis, le protocole est mis sous la forme d'un ensemble de règles grâce à la plaquette et à l'aide précieuse d'une spécialiste en ergonomie qui est également médecin et qui joue le rôle d'« interface » avec les experts. Enfin, le travail de traduction de ces règles en classes est effectué.

3.4 Lien avec un travail proche : RÉSYN

L'utilisation de la classification dure pour modéliser le raisonnement protocolaire s'est fortement inspirée du système RÉSYN [Vismara, 1995] dont le domaine d'application est l'aide à la

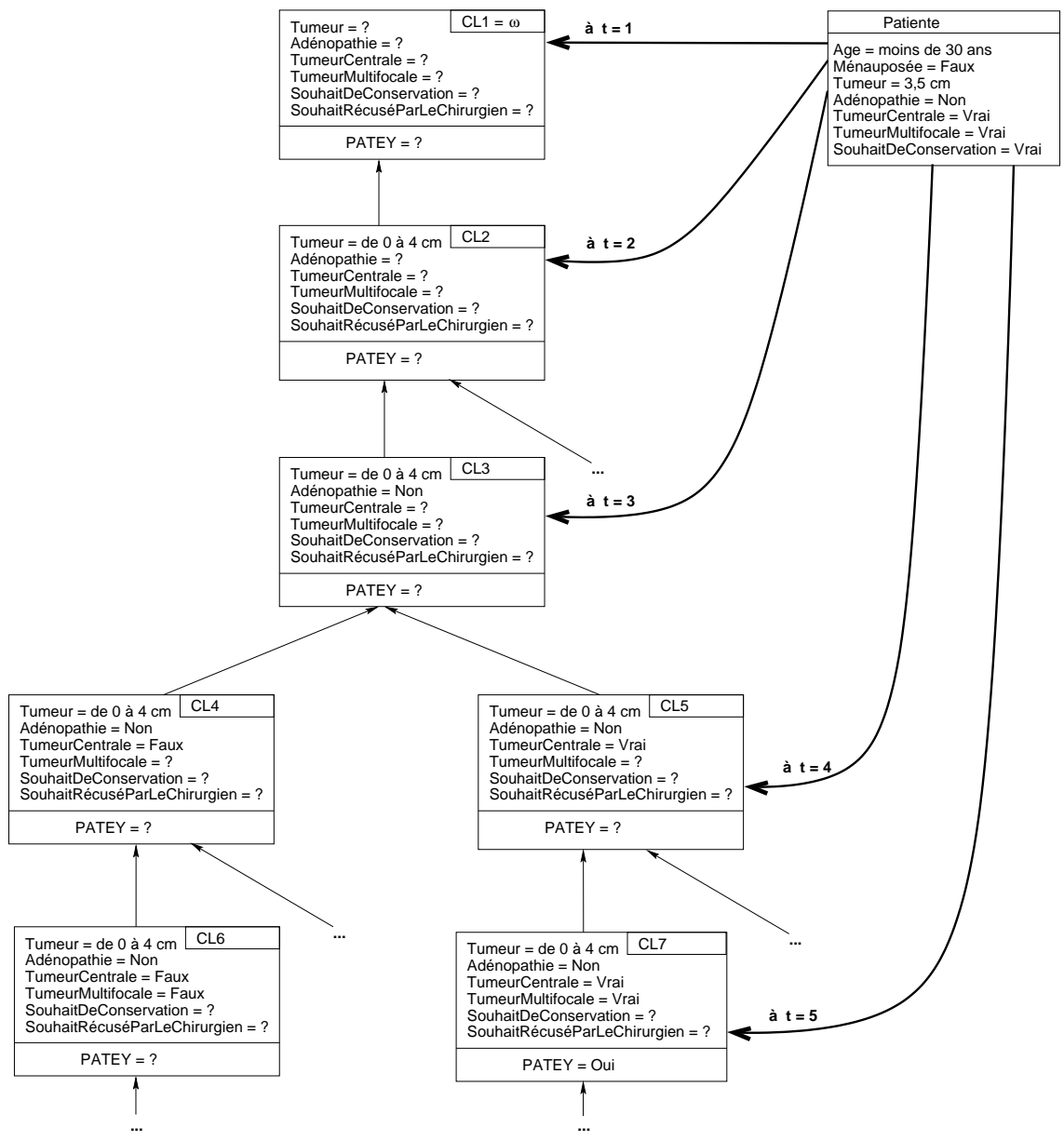


FIG. 1 – Exemple de classification dure dans une partie de la hiérarchie (notons que les classes représentées sont très simplifiées par rapport aux classes effectivement représentées dans CASIMIR).

planification de synthèse en chimie organique. Le système RÉSYN a été implanté dans le cadre du groupement de recherches du CNRS « Traitement Informatique de la Connaissance en Chimie Organique ». Un autre travail effectué dans le cadre de ce GDR était l'utilisation du raisonnement à partir de cas pour le même domaine d'application. Le système RÉSYN/RÀPC [Lieber et Napoli, 1996; Lieber, 1997] peut être vu comme une extension de RÉSYN, la différence réside essentiellement dans le fait que la classification dure est suivie d'une classification élastique.

4 Raisonnement non protocolaire et classification élastique

Pour environ 30% des patients, un raisonnement non protocolaire est nécessaire. Cela peut arriver dans deux types de situations :

- (S1) Soit le protocole ne s'applique pas. Cela signifie qu'il n'y a pas dans la hiérarchie de classe CL plus générale que CL_{patient} .
- (S2) Soit le protocole s'applique mais le traitement associé pose problème. Par exemple, si le traitement propose une ovariolyse (ablation des ovaires) et que la patiente n'a que 30 ans et aucun enfant, le cancérologue peut avoir des scrupules à effectuer ce traitement.

Dans les deux situations, les classes du protocoles sont utilisées, mais de façon moins absolues que lors du raisonnement protocolaire. Plus précisément ces classes (et les traitements associés) sont *adaptées* au patient. Elles peuvent donc être considérées comme des cas, au sens du raisonnement à partir de cas [Riesbeck et Schank, 1989] :

Remémoration :

la classe CL *la plus proche* de CL_{patient} est recherchée dans l'ensemble des classes ;

Adaptation :

le traitement associé à cette classe est adaptée pour proposer un traitement au patient.

Nous n'étudierons ici que les situations de type (S1) et envisageons de traiter les situations de type (S2) de façon similaire.

Le principe de la remémoration par classification élastique est le suivant : on cherche une classe CL telle que « à quelques modifications près sur CL et CL_{patient} », CL soit plus générale que CL_{patient} . Plus formellement, on cherche CL, φ et ψ telles que $\varphi(CL) \succcurlyeq \psi(CL_{\text{patient}})$. φ et ψ sont appelées *fonctions de modification de classes*. Le choix (ou la construction) de φ et ψ n'est pas indifférent : il est lié aux connaissances d'adaptation dont on dispose. En effet, si on note $Ttt(CL)$ le traitement associé à CL, on veut pouvoir réaliser l'adaptation de ce traitement en un traitement $Ttt(CL_{\text{patient}})$ destiné au patient, en passant par les classes intermédiaires $\varphi(CL)$ et $\psi(CL_{\text{patient}})$. Plus précisément, l'adaptation se fera selon le schéma suivant (appelé *chemin de similarité dans [Lieber, 1997]*) :

$$CL \xrightarrow{\varphi} \varphi(CL) \succcurlyeq \psi(CL_{\text{patient}}) \xleftarrow{\psi} CL_{\text{patient}}$$

Ce qui signifie que les étapes de l'adaptation sont :

- (φ) Adaptation de $Ttt(CL)$ en un traitement $Ttt(\varphi(CL))$ de $\varphi(CL)$;
- (\succcurlyeq) $Ttt(\psi(CL_{\text{patient}})) := Ttt(\varphi(CL))$
(comme $\varphi(CL) \succcurlyeq \psi(CL_{\text{patient}})$, le traitement associé à $\varphi(CL)$ est copié et associé à $\psi(CL_{\text{patient}})$);
- (ψ) Adaptation de $Ttt(\psi(CL_{\text{patient}}))$ en un traitement $Ttt(CL_{\text{patient}})$ de CL_{patient} .

Il peut y avoir plusieurs triplets (CL, φ, ψ) permettant de faire l'adaptation ci-dessus tout comme, pour la classification dure, il pouvait y avoir plusieurs classes CL plus générales que CL_{patient} . Le

problème de la sélection du « meilleur » triplet n'a pas encore été étudié pour CASIMIR. Pour l'application RÉSYN/RÀPC évoquée plus haut, le choix est effectué grâce à une fonction coût qui aux fonctions de modification associe les nombres réels positifs $\text{coût}(\varphi)$ et $\text{coût}(\psi)$. Le triplet sélectionné est celui qui minimise $d(\text{CL}, \text{CL}_{\text{patient}}) = \text{coût}(\varphi) + \text{coût}(\psi)$ (1). Pour cette application, φ est une fonction de généralisation de graphes moléculaires — $\text{CL} \preceq \varphi(\text{CL})$ — construite à l'aide de règles de généralisations et ψ est lié à des connaissances sur les réactions chimiques (pour plus d'information, voir [Lieber et Napoli, 1996] ou [Lieber, 1997]).

La classification élastique s'appuie donc sur des *connaissances d'adaptation* qui permettent de réaliser les étapes (φ) et (ψ) ci-dessus. Nous aborderons le problème de l'acquisition de ces connaissances au paragraphe suivant.

Une remarque importante doit être faite ici. Un tel raisonnement non protocolaire, puisqu'il est incertain, se doit d'être effectué en collaboration avec les oncologues, dans le cadre du comité de décision thérapeutique. Cela suppose que le raisonnement non protocolaire de CASIMIR soit présenté de façon intelligible pour les médecins, de façon à pouvoir facilement être remis en cause.

5 Évolution du protocole

Les connaissances utilisées pour le raisonnement protocolaire sont représentées par la hiérarchie des classes de patients et les traitements correspondant. Les connaissances utilisées pour le raisonnement non protocolaire sont représentées par cette hiérarchie et par les connaissances d'adaptation. Ces dernières peuvent être considérées comme des *règles correctives* du protocole. Parfois, ces connaissances s'avèrent insuffisantes pour proposer un traitement ou alors, ces connaissances donnent un traitement inadapté compte tenu des connaissances actuelles en oncologie du sein (c'est un domaine en constante évolution). Le raisonnement aboutit alors à un échec. Pour faire évoluer les connaissances, on cherche à faire un *apprentissage à partir d'échecs* qui consiste à analyser l'échec et à acquérir, avec l'aide des experts du domaine de la cancérologie du sein, de nouvelles connaissances qui permettent d'éviter à l'avenir cet échec et certains échecs analogues [Boussu, 1998]. Cette analyse est effectuée en interaction avec l'expert.

Cet apprentissage à partir d'échecs permet d'affiner progressivement les connaissances d'adaptation. Quand celles-ci sont jugées suffisamment stables, on peut proposer de les intégrer dans les connaissances protocolaires, ce qui conduit à une correction de la hiérarchie et des traitements qui y sont associés. L'apprentissage à partir d'échecs permet ainsi de faire évoluer le protocole.

6 Conclusion

L'application CASIMIR présentée dans cet article est en cours de développement. Au stade actuel, les classes mises en jeu pour le raisonnement protocolaire sont en cours d'implantation. Ce que nous avons décrit dans cet article est essentiellement la ligne de conduite que nous avons choisie de suivre. Elle est inspirée des applications RÉSYN et RÉSYN/RÀPC d'aide à la planification de synthèse en chimie organique. Cependant, l'accent est davantage mis dans CASIMIR sur le problème de l'acquisition des connaissances. La première perspective est de poursuivre le travail en cours : finir l'acquisition des connaissances protocolaires, implantation des classes du protocole, implantation de la classification dure, précision de certains points concernant la classification élastique et l'apprentissage à partir d'échecs, implantation de la classification élastique, implantation

1. d est similaire à une distance d'éditions (voir par exemple [Bunke et Messmer, 1994]). Pour RÉSYN/RÀPC, c'est par abus de langage qu'on parle de distance, en effet, d n'est, en particulier, pas symétrique pour l'espace des fonctions de modifications φ et ψ associées à l'application RÉSYN/RÀPC.