



" Modèle de recherche de trajectoires patients dans le cadre de pathologies infectieuses et tropicales : exemple du paludisme "

Claude Petit - Claude Dussart Centre Européen de Santé Humanitaire de Lyon

Claude Petit - Claude Dussart
Centre Européen de Santé Humanitaire de Lyon
Laboratoire LASS UMR 5823 CNRS
Université Claude Bernard - Lyon 1
43, Bd. du 11 Novembre 1918
69622 Villeurbanne
c-petit@univ-lyon1.fr - dussart.claude@wanadoo.fr

RESUME

Le choix d'une stratégie thérapeutique est soumis, pour une pathologie donnée, à une grande variabilité inter-individuelle : chaque prescripteur élabore son choix, non seulement à partir de ses connaissances de la pathologie et des traitements adaptés, mais aussi en fonction de son expérience, de sa formation d'origine...

Le but de notre travail est de détecter et d'analyser des trajectoires patients (qui correspondent au suivi dans le temps des traitements adaptés à un patient donné). La méthode d'analyse met en œuvre le concept d'Algorithme Génétique.

Notre propos est illustré par l'exemple du Paludisme. A partir d'une base de 52 cas patients, l'Algorithme Génétique déployé a mis en évidence 3 grandes tendances de stratégies thérapeutiques. L'approche originale que nous avons développé s'attache alors à apporter un nouvel éclairage des problèmes de choix thérapeutique et s'inscrit dans le cadre de la planification des soins et de l'aide à la décision.

RESUMEN

La opción de una estrategia terapéutica es sometida, para una patología dada, tiene una variabilidad grande inter-individual: cada prescriptor elabora su opción, no sólo del conocimiento de la patología y los tratamientos adaptados, pero también según su experiencia, a su formación, la educación de origen ...

El objetivo de nuestro trabajo es de descubrir y analizar a pacientes de carreras (que se corresponde a la continuación en el tiempo de tratamientos adaptados a un paciente dado). El método analítico usa el concepto de Algoritmo Genético.

Nuestro comentario es ilustrado por el ejemplo de la Malaria. De una base de 52 casos pacientes, extensión el Algoritmo Genético puesto en la evidencia 3 tendencias de estrategias terapéuticas. El acercamiento original que desarrollamos tentativas luego para traer una iluminación nueva de los problemas de opción terapéutica y juntas dentro del marco de la planificación económica de cuidado y la ayuda la decisión.

PROBLEMATIQUE

Dans le domaine de la santé, le choix d'une stratégie thérapeutique s'avère souvent difficile. En réalité, il existe un décalage entre les données scientifiques et la réalité des pratiques. Un prescripteur est influencé dans son choix par sa formation d'origine, sa sensibilité, sa connaissance de la pathologie et des médicaments, et par son expérience.

On peut alors supposer que, pour une pathologie donnée, différents traitements peuvent être mis en œuvre. Un traitement administré à un patient donné, suivi dans le temps, représente une trajectoire patient pour une pathologie identifiée. Notre travail se propose de détecter et de représenter des

trajectoires patients afin de déterminer la ou les trajectoires types issues de l'expérimentation contrôlée ou de la pratique. La méthode développée s'appuie sur le concept d'Algorithme Génétique (AG). Il s'agit d'un algorithme capable de résoudre des problèmes dont on ne connaît pas de méthode de résolution ou dont la solution exacte est trop complexe pour être calculée en un temps raisonnable.

Nous illustrerons notre propos par l'exemple du Paludisme. En effet, l'approche thérapeutique du paludisme a été profondément modifiée au cours de la dernière décennie. On assiste à l'extension de la chloroquino-résistance de *Plasmodium falciparum*, à l'apparition dans de nombreux pays de multirésistances et à des diminutions de sensibilité vis à vis de la quinine qui reste l'antimalarique de dernière intention. Dans le même temps, des molécules largement prescrites se révèlent délétères, induisant parfois des accidents cardiaques et/ou neurologiques. Toutes ces raisons expliquent l'engouement actuel pour de nouvelles molécules et la nécessité d'établir de meilleures règles d'utilisation des antimalariques.

Les informations émergentes permettront d'améliorer la qualité des soins en générant une information robuste et une aide à la décision susceptibles de modifier et d'améliorer les pratiques professionnelles.

1. MATERIELS ET METHODES

1.1. MATERIELS

Trajectoire patient

La modélisation des trajectoires patients repose sur la définition générale suivante : **une trajectoire patient est une succession dans le temps d'événements relatifs à un patient ou à un groupe de patients, relevant d'un problème ayant provoqué cette succession d'événements.**

Considérons les éléments suivants :

- la population des patients. Nous posons $\|\Omega\| = N$.
- P l'ensemble des différentes pathologies pouvant affecter ces patients. Nous posons $\|P\| = n$.
- M l'ensemble des médicaments pouvant être prescrits. Nous posons $\|M\| = m$.
- le temps représenté par l'indice t.

Soit $i, i=1, \dots, k, \Omega_i \subset \Omega$, l'ensemble des patients atteints de la pathologie i. De manière évidente, Ω_i possède les propriétés suivantes :

$$\forall i, \Omega_i \neq \emptyset$$

$$\Omega = \bigcup_{i=1}^n \Omega_i$$

En revanche, on voit que :

$$\Omega_i \cap \Omega_j \neq \emptyset$$

La famille $\{\Omega_i, i = 1, \dots, n\}$ forme donc un recouvrement de Ω

Considérons l'application $c_i(\cdot)$ de $\Omega_i \times \mathbb{R}^+$ dans $P(M)$, ensemble des parties de M, qui à tout patient ω de Ω_i associe l'ensemble $c_i(\omega, t)$ des médicaments prescrits au patient à l'instant t.

Cette application $ci(.)$ est appelée la multiapplication " prescrite instantanée " de Ω_i dans M . Le problème est donc de connaître $\{ ci(\omega, t), t \in \mathbb{R}^+, \forall i = 1 \dots p, \forall \omega \in \Omega_i$.

Base de données

Pour ce travail, la base de données doit être constituée d'un grand nombre de dossiers patients relevant d'une même pathologie. Elle est articulée autour de 20 variables. Ces variables permettent de décrire de manière la plus synthétique et précise possible le scénario suivant : un patient se rend chez un médecin à une date donnée, la nature de la pathologie est révélée par l'agrégation des résultats des examens cliniques et paracliniques (biochimiques, biologiques et radiologiques). Selon la pathologie et le contexte patient (age, poids, sexe...), un traitement est mis en œuvre. Le traitement est constitué d'un ensemble de médicaments. Dans le cas de pathologies complexes, leur évolution sera suivie au cours de consultations successives.

On distingue les variables actives et passives. Les variables actives sont celles qui participent à la recherche de la ou des trajectoires optimales. Les variables passives seront ensuite utiles pour affiner la recherche.

Variables actives

- Date de prescription
- Traitement administré 1
- Indice d'évolution 2

Variables passives

- Nom du patient
- Age du patient
- Sexe du patient
- Poids du patient
- Catégorie socioprofessionnelle du patient
- Nom du médecin prescripteur
- Cabinet ou service hospitalier de rattachement du prescripteur
- Spécialité (qualification) du prescripteur
- Diagnostique 3
- Indicateur de gravité 4
- Types d'examens cliniques pratiqués
- Résultats des examens cliniques
- Types d'examens complémentaires
- Résultats des examens complémentaires
- Type de bilan biologique et/ou biochimique
- Résultats des examens biologiques et/ou biochimiques
- Notes complémentaires

1 : les traitements sont constitués d'un ensemble de médicaments qui permettent de mettre en œuvre une thérapeutique pour une pathologie donnée

2 : l'indice d'évolution reflète la " distance " entre la situation du patient et celle d'un témoin (absence de pathologie).

3 : les diagnostics permettent la catégorisation d'une pathologie donnée

4 : l'indicateur de gravité renseigne sur le degré de gravité de la pathologie.

Codification

Le **diagnostique** est codifié en utilisant une classification définie au préalable. Les classifications permettent d'organiser les données de manière systématique. C'est un moyen commode pour représenter et codifier les concepts médicaux. De nombreuses classifications existent : certaines sont propres à des organisations ou à des services, d'autres ont une vocation internationale (telle la classification internationale des maladies).

Le vocabulaire du domaine médical est riche et imprécis. Pour un terme donné, l'ensemble de la communauté médicale ne reconnaît pas la même définition. Une maladie peut alors être désignée par

des noms ou des expressions différents (synonymie) ; le même terme peut avoir plusieurs significations suivant le contexte (polysémie). D'où l'importance de règles de codage précises et sans ambiguïté.

Le traitement est constitué d'un ensemble de médicaments administrés au patient.

Il est constitué de

- médicaments à visée curative : A, B ou C
- médicaments complémentaires (effet de synergie) : D ou E
- médicaments divers (anti-émétiques, antibiotiques) : F, G ou I

Pour la base de cas, le traitement sera constitué de quatre médicaments : un curatif, un complémentaire et deux divers. L'absence de l'un d'eux sera notée 0 (zéro).

Indice d'évolution

Cet indice varie entre 0 et 100. Plus cet indice tend vers 100, plus le patient se rapproche du témoin.

Un indice à 100 signifie que le patient est guérit.

Le suivi de l'efficacité d'un traitement est assuré par :

- Les examens cliniques (inspection, palpation, auscultation...)
- Les examens paracliniques (biologiques, biochimiques, radiologiques...)

A chaque consultation, les examens cliniques et paracliniques sont mis en œuvre. Les variations et/ou la distance entre la valeur mesurée et la valeur normale indiquent (avec un poids plus ou moins important selon chaque examen) l'évolution de la pathologie (i.e. l'efficacité relative du traitement). Pour une pathologie donnée, chaque examen (et sa variation afférente) influe ainsi sur l'indice d'évolution de la pathologie avec un poids qui lui est propre.

L'affectation des poids peut-être défini dans le cadre d'une conférence de consensus. En effet, rappelons que la conférence de consensus réalise une synthèse des connaissances sur un sujet. Elle fournit une aide à la décision médicale et contribue à clarifier des controverses médicales. Cette méthode de jugement relève ainsi d'un schéma empruntant des éléments de trois modèles différents : le modèle judiciaire où des témoins (experts et publics) sont écoutés par un jury impartial et éclairé ; la réunion scientifique au cours de laquelle des experts exposent et discutent leurs travaux ; le débat démocratique où chaque personne peut exprimer son point de vue.

1.2. ALGORITHMES GENETIQUES

Présentation

Souvent, lorsque l'on doit résoudre un problème d'aide à la décision, on se trouve confronté à plusieurs critères contradictoires qui varient en sens opposé. Les personnes, qui modélisent les problèmes, se ramènent à des problèmes mono critères parce qu'il est trop difficile de travailler avec plusieurs critères. Ils accordent des poids aux critères, travaillent sur une somme pondérée de critères ou hiérarchisent l'importance des critères.

Depuis environ 1975, toute une théorie a été développée sur l'approche multicritère en aide à la décision. Elle s'intéresse essentiellement à la manière de monter des procédures (généralement interactive de décision) qui permettent de promener le décideur dans l'espace des décisions en se limitant à des solutions "intéressantes" (appelées dans la littérature "solutions efficaces" ou "optima de Pareto") et à l'amener à prendre sa décision finale.

L'algorithme génétique permet d'obtenir des solutions à un problème n'ayant pas de méthode de résolution décrite précisément, ou dont la solution exacte, si elle est connue, est trop compliquée pour être calculée en un temps raisonnable. C'est notamment le cas quand des contraintes multiples et complexes et parfois même en partie contradictoires doivent être satisfaites simultanément.

Principe de base

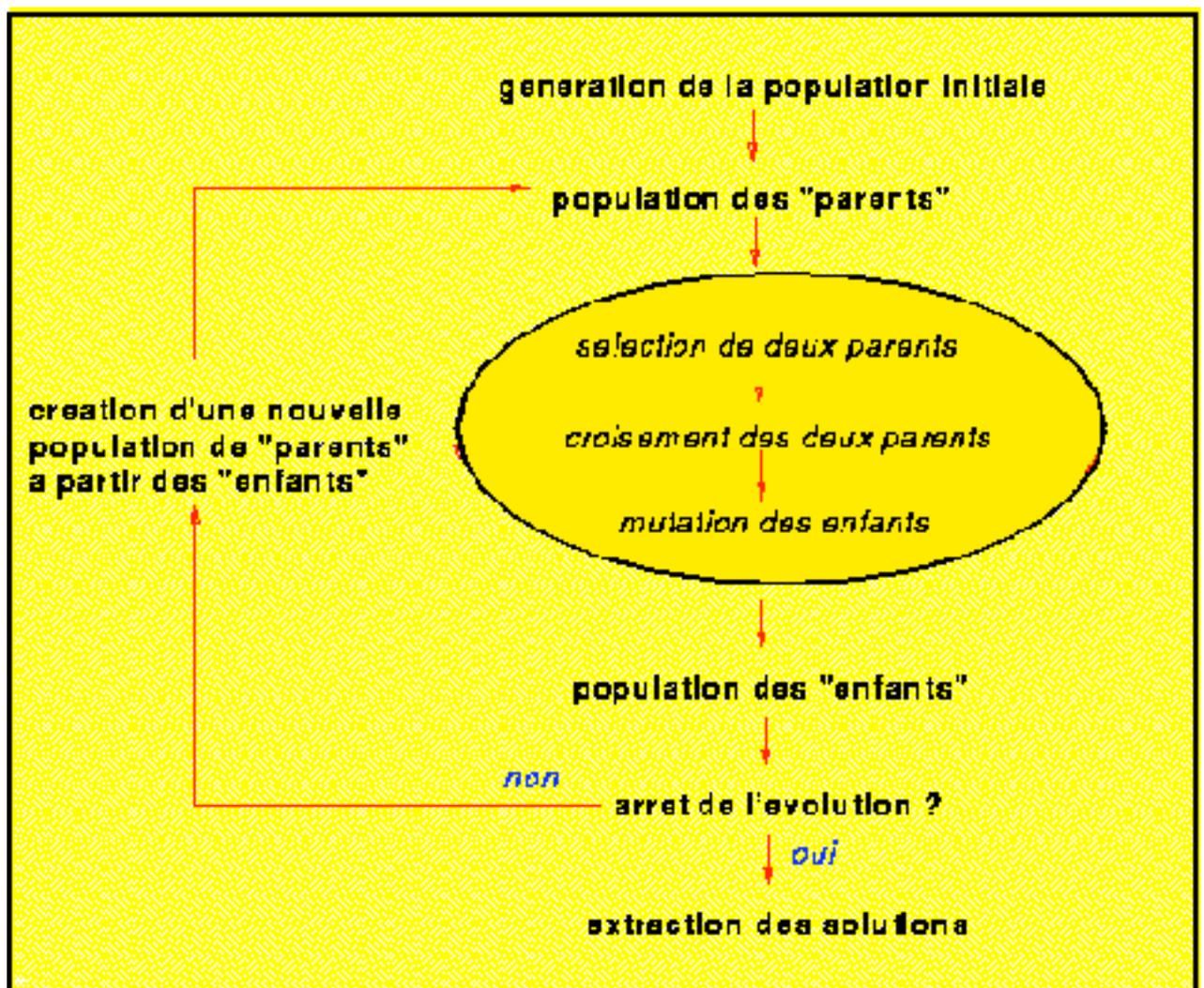
Selon l'algorithme génétique, de nombreuses solutions, plus ou moins bonnes, au problème donné sont créées au hasard, selon une forme définie à l'avance : itinéraire, emploi du temps, base de règles de décision, grille de score, réseau de neurones, etc.

Ces solutions sont représentées chacune par une chaîne de 0 et de 1 ou "chromosomes", qui sont alors soumis à une imitation de l'évolution des espèces : mutations et reproduction par hybridation. En favorisant la survie des plus "aptés" (les solutions les plus correctes), on provoque l'apparition d'hybrides meilleurs que chacun de leurs parents. Une "fonction de décodage" sert à "traduire" la représentation sous forme de chromosomes 0-1 en la forme qui nous intéresse.

La population initiale donne ainsi naissance à des générations successives, mutés et hybridés à partir de leurs "parents". Le mécanisme favorise les éléments les plus aptes ("pression de l'évolution") de sorte que les générations successives soient de plus en plus adaptées à la résolution du problème. Ce mécanisme surprenant a été rigoureusement validé mathématiquement (notamment par John KOZA dans les années 70).

Le mécanisme d'évolution et de sélection est indépendant du problème à résoudre : seules varient la fonction qui décode le génotype en une solution possible (tout décodage peut être utilisé, de préférence le plus simple possible) et celle qui évalue la justesse de la solution. Cette technique est d'application générale.

Ainsi, les Algorithmes Génétiques, basés sur la théorie de l'évolution de Darwin, consistent à faire évoluer une population de dispositifs à l'aide de différents opérateurs : sélection, croisements, mutations.



Conclusion

Pour mettre en œuvre un AG, il faut définir :

1. une représentation chromosomique des solutions au problème,
2. une méthode de création de la population initiale des solutions
3. une fonction d'évaluation qui joue le rôle de l'environnement,
4. des opérateurs génétiques qui modifient la composition des chromosomes,
5. un certain nombre de paramètres :
 - taille de populations,
 - probabilité d'application des opérateurs génétiques,
 - critère d'arrêt,
 - etc....

2. EXEMPLE DU PALUDISME

2.1. PRESENTATION

Le paludisme est la maladie parasitaire la plus répandue. 40 % de la population mondiale vit en zone endémique. Le paludisme touche, à titre d'exemple, 3 à 4 mille personnes en France suite à un séjour dans ces zones. Il se répartit en trois zones de danger inégal en fonction de la résistance à la chloroquine : zone A à risque limité (pas de résistance), zone B (résistance partielle), zone C (résistance fréquente ou multirésistance aux antimalariques).

La lutte contre le paludisme s'effectue selon plusieurs axes : la lutte antivectorielle, la chimioprophylaxie et le traitement curatif. L'apparition de chimiorésistance du Plasmodium a nécessité le développement de nombreux traitements antipaludiques à base d'un ou plusieurs antipaludéens.

Classification des antipaludéens selon leur mode d'action :

- Les schizonticides érythrocytaires
- Les schizonticides érythrocytaires et tissulaires
- Les gamétocytocides

On différencie les primo-infections, généralement plus dangereuses, des accès de fièvre périodique. Des complications apparaissent parfois comme le paludisme viscéral évolutif, le neuropaludisme, la fièvre bilieuse hémoglobinurique et certaines néphrites.

2.2. RESULTATS ET DISCUSSION

Notre travail s'est appuyé sur une base de cas de 52 dossiers patients.

L'algorithme génétique travaille à partir d'un ensemble d'éléments choisis au hasard dans un espace de solutions potentielles. Sa connaissance du problème à résoudre se résume à une fonction d'évaluation qui mesure la qualité d'un élément en tant que solution au problème. *La fonction d'évaluation sera ici l'indice d'évolution.* La qualité du maximum initial est bien entendu très médiocre puisqu'il a été choisi au hasard. Cependant, l'algorithme génétique va tirer parti des meilleurs éléments à sa disposition pour fabriquer de toutes pièces un nouvel ensemble d'éléments répondant en moyenne de mieux en mieux au problème.

Pour passer de l'ancien au nouvel ensemble, les éléments sont sélectionnés aléatoirement et proportionnellement à leur qualité en tant que solution au problème. *Ainsi, le traitement avec le meilleur indicateur d'évolution aura plus de chances d'être sélectionné que les autres éléments. S'il est retenu, il contribuera à améliorer les éléments du nouvel ensemble. Dans le cas contraire, il est irrémédiablement perdu pour les ensembles futurs.*

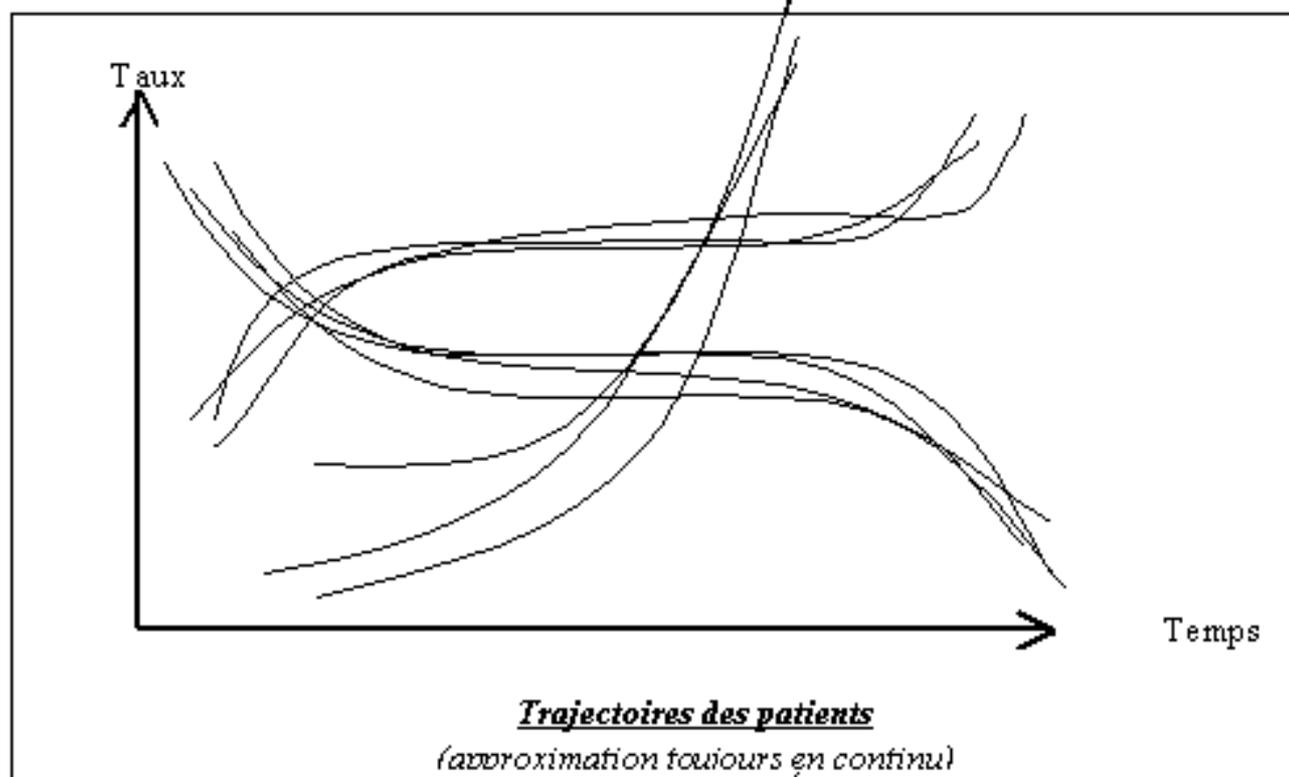
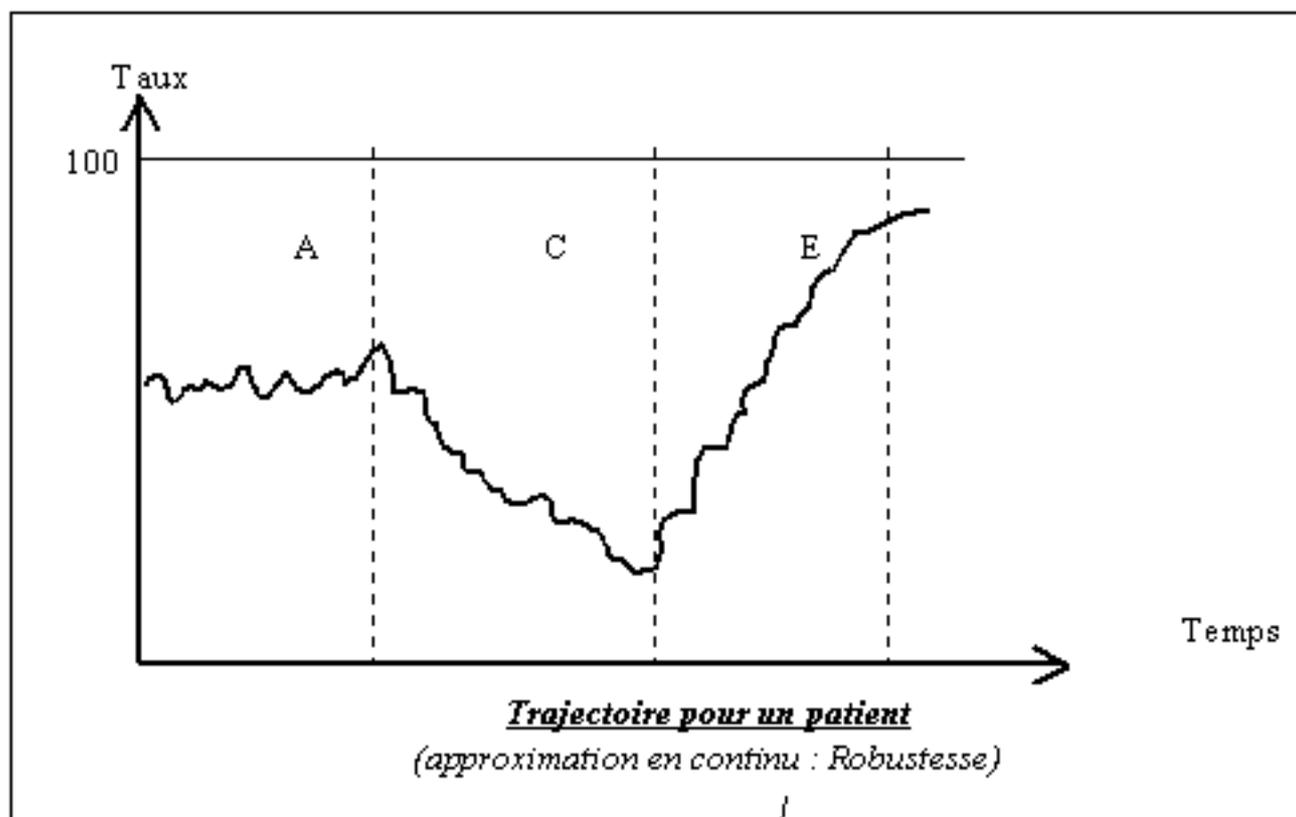
Les éléments sélectionnés sont ensuite combinés en de nouveaux éléments grâce à des méthodes inspirées de phénomènes naturels, comme le croisement génétique ou la mutation. *Les méthodes ne seront pas détaillées dans ce travail. Il suffit d'admettre que nous obtenons un nouvel ensemble parmi lequel nous recherchons le nouvel optimum.*

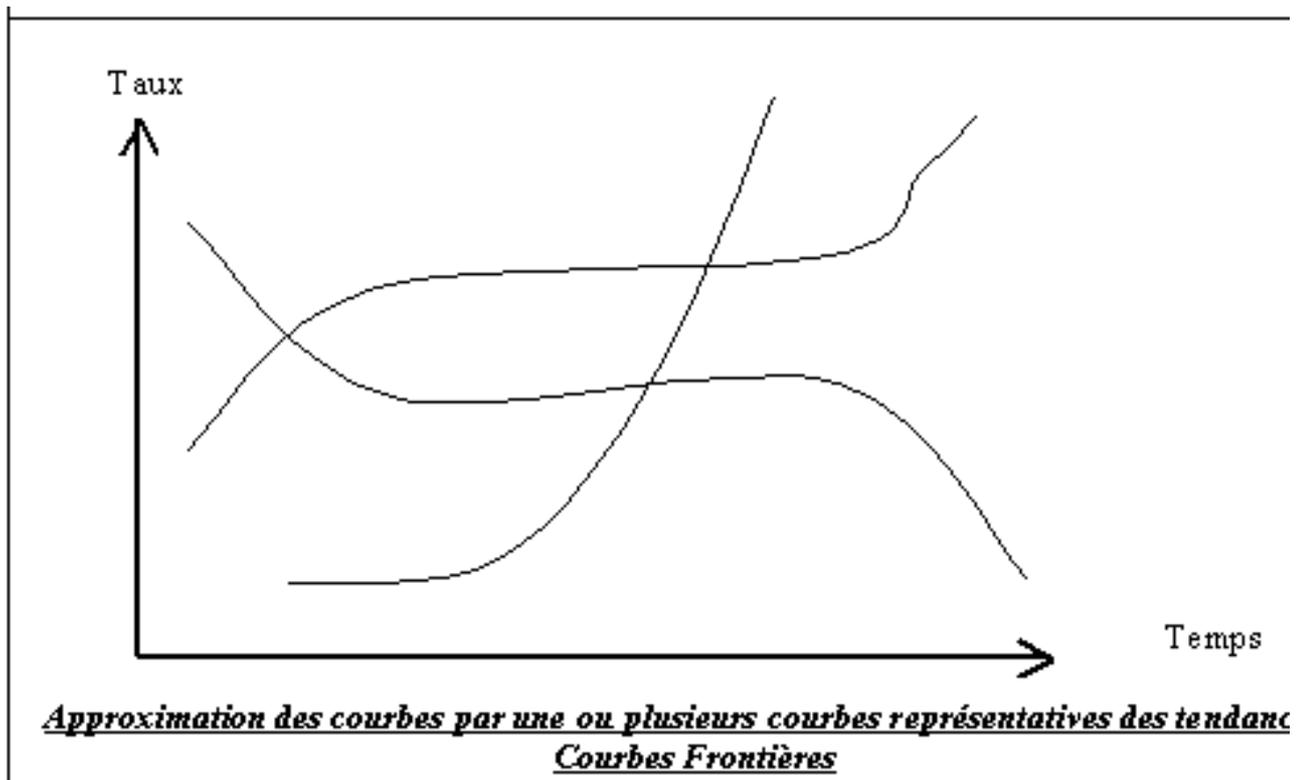
L'algorithme génétique travaille ensuite par générations successives jusqu'à un critère d'arrêt :

- soit le temps imparti est écoulé,
- soit le meilleur élément de la dernière génération atteint un seuil de qualité fixé au départ...

Il rend alors une solution approchée et acceptable en un temps bien inférieur à celui qu'aurait demandé la recherche de la solution exacte du problème.

Les résultats sont : (pour une période donnée ; soit A, C et E les traitements administrés.)





La programmation génétique définit une courbe polynomiale, caractéristique d'une tendance (c'est à dire d'une stratégie thérapeutique optimisée). Les tendances sont ainsi approximées par une fonction.

A partir de notre travail, nous avons mis en évidence trois grandes courbes frontières. Chacune représente une stratégie thérapeutique type. Nous ne développerons pas l'analyse détaillée des courbes frontières. Leur justification thérapeutique n'est pas de circonstance. Il faut retenir de cet exemple qu'il est possible de déterminer les grandes tendances de choix thérapeutiques optimisés, caractéristiques d'une pathologie donnée, à partir de grandes bases de données. Ces tendances types sont représentées sous forme de courbes frontières.

On parle de tendance car on se place dans un domaine continu et non discret. L'approximation en continu assure alors la robustesse de la technique. Les points " anormaux " ne sont pas pris en compte. Cet effet de lissage permet d'appréhender le problème de manière plus aisée.

3. DISCUSSION GENERALE

De nombreux travaux théoriques ont été effectués pour évaluer l'efficacité des algorithmes génétiques. Il a été démontré mathématiquement que pour une population de n individus, le nombre de groupes de gènes voisins couverts d'une manière sous-jacente est de l'ordre de n puissance 3 ? En d'autres termes, pour un problème nécessitant une recherche parmi un million de possibilités, il suffira de faire évoluer cent individus.

Il s'agit d'une approche un peu brutale, nécessitant une grande puissance de calcul, mais présentant l'immense avantage pratique de fournir des solutions pas trop éloignées de l'optimal, même si l'on ne connaît pas de méthode de résolution. Contrairement à la recherche opérationnelle, l'algorithme génétique n'exige aucune connaissance de la manière dont résoudre le problème; il est seulement nécessaire de pouvoir évaluer la qualité d'une solution. Il est également léger à mettre en œuvre (le "moteur" est commun, il y a moins de programmation spécifique au problème à faire).

	Recherche opérationnelle (approche ad hoc, analytique, spécifique)	Approche génétique
Rapidité	Selon solution, parfois bonne	Bonne à très bonne
Performance	Selon solution	Bonne

Compréhension au problème	Nécessaire	Non nécessaire
Travail humain	Très variable : de quelques jours à quelques thèses	Faible : quelques jours ou semaines
Applicabilité	Faible : la plupart des problèmes intéressants n'ont pas d'expression mathématique exploitable	Générale
Étapes intermédiaires	Ne sont pas des solutions : il faut attendre la fin des calculs	Sont des solutions : le processus peut être interrompu à tout moment.

Les applications industrielles des algorithmes génétiques sont nombreuses : diagnostic de pannes, configuration optimale de pipe-line, établissement de prévisions dans les secteurs industriels, commerciaux et bancaires, gestion du trafic dans de grandes villes...

Son utilisation dans le domaine de la santé, et tout particulièrement celui de la planification des soins, est novatrice et prometteuse. Elle s'inscrit dans une recherche d'aide à la décision.

CONCLUSION

Le but de ce travail est de formuler des propositions concrètes, réellement utiles et utilisables par les acteurs de la santé.

Ils sont en particulier utilisés pour les problèmes d'optimisation comportant de multiples paramètres et des objectifs multiples.

L'outil utilisé, un Algorithme Génétique, a montré son aptitude à donner des solutions concrètes dans une situation complexe. On observera sa robustesse et sa capacité à donner des résultats faciles à interpréter.

La représentation de courbes types qui exprime un système d'ensemble de trajectoires patients est déterminée de façon originale en utilisant des courbes frontières. L'Algorithme Génétique définit automatiquement les paramètres de chaque courbe type.

BIBLIOGRAPHIE

[1] Charmot G., Coulaud J.P. Les formes atypiques du paludisme à Plasmodium falciparum chimio-résistant observées en France. Méd Mal Infectieuses 1988, 2 :52-56.

[2] Giraud A. Jolly D. Origines et définition des conférences de consensus. In : A Giraud, D Jolly. Le consensus en médecine : analyse et bilan des conférences de consensus dans le monde. Paris. Doin, 1991 : 61-67.

[3] Goldberg D.E. Genetic algorithms in search, optimization and machine learning. Addison Wesley. 1989.

[4] Holland J.H. Adaptation in natural and artificial systems. MIT Press. 1992.

[5] Lebras J. Actualités et perspectives en thérapeutique antipaludique. Med Trop. 1995, 55 : 7-13.

[6] Luzzi G.A. Adverse effects of anti-malarials, an update. Drug Saf 1993, 8 :295-311.

[7] Marsala C. Apprentissage et acquisition des connaissances ? Thèse de doctorat. Paris 6. 1998.

[8] Saporta. Probabilités, analyse des données et statistique. Technip. 1990.