

Revue Internationale de

ISSN 0980-1472

systemique

CONNAISSANCES IMPLICITES
ET
CONNAISSANCES EXPLICITES

Vol. 6, N° 1-2, 1992

afcet

DUNOD

AFSCET

Revue Internationale de
systemique

Revue
Internationale
de Sytémique

volume 06, numéro 1-2, pages 115 - 125, 1992

Un protocole d'élicitation de l'expertise

Laurent Bochereau
Paul Bourguine
Guillaume Deffuant

Numérisation Afscet, août 2017.



UN PROTOCOLE D'ÉLICITATION DE L'EXPERTISE

L. BOCHEREAU, P. BOURGINE, G. DEFFUANT
CEMAGREF ¹

Résumé

L'élicitation directe des règles auprès d'un expert s'avère difficile. Nous proposons un protocole d'élicitation des règles, qui se déroule en deux temps : on cherche d'abord à obtenir un réseau connexionniste, fonctionnellement proche du savoir-faire de l'expert, représenté par une base d'exemples; le classificateur connexionniste est à son tour équivalent à un ensemble de clauses.

On présente ensuite un exemple d'extraction d'un ensemble de clauses équivalentes. L'examen de la microstructure des clauses conduit à conjecturer qu'une bonne précision ne peut être obtenue qu'à partir d'un nombre exponentiellement croissant de clauses. Si cette conjecture s'avérait juste, cela permettrait de mieux comprendre les difficultés de l'élicitation directe des règles et renforcerait l'intérêt des protocoles d'élicitation à partir d'exemples.

Abstract

The direct elicitation of expertise rules has proved to be a difficult issue. We suggest a rule elicitation protocol that consists of two steps: firstly, we construct a connectionist network trained on a data base representing the expert's know-how; secondly, we construct the clauses set equivalent to the connectionist classifier.

We present an example of the extraction of such an equivalent clauses set. The analysis of clause microstructure leads us to conjecture that a high precision is obtained only through an exponentially increasing set of clauses. If this conjecture proved to be right, we could better understand why direct elicitation is difficult; this would also reinforce the interest of such elicitation protocols using examples.

Les besoins de la communication ne conduisent pas à expliciter toutes les connaissances sous une forme logique : les connaissances de sens commun,

1. Parc de Tourvoie, 92185 Antony Cedex.

partagées, n'ont pas besoin d'être explicitées dans la communication. Et elles forment, de loin, la plus grande part des connaissances. Les connaissances non partagées sont liées à une expertise particulière; l'activité scientifique comme les rapports sociaux dans la société contemporaine conduisent de plus en plus à éliciter les connaissances expertes.

Le but de cet article est de proposer un protocole d'élicitation de l'expertise. Il s'agit d'une question très difficile : Hubert Dreyfus (1972) considérait même la tâche comme impossible, puisque, selon lui, un expert ne raisonne pas avec des règles et donc un système expert ne peut atteindre que le niveau d'un novice; la position de Herbert Simon (1988) n'est guère plus confortable, puisqu'il considère qu'un vrai domaine d'expertise, tel celui d'un maître international aux échecs, est de l'ordre de plusieurs centaines de milliers de règles.

Il n'est pas question avec cet article d'apporter une réponse définitive à cette question, mais seulement d'ouvrir une voie de recherche, qui permette d'éliciter plusieurs milliers de règles, représentant de façon approchée le savoir-faire d'un expert. Parce qu'il ne nous semble guère possible d'obtenir une telle quantité de règles directement auprès de l'expert, nous avons choisi d'approcher son savoir-faire par apprentissage à l'aide d'une base d'exemples.

On présentera tout d'abord le type d'expertise considérée et le principe du protocole. Un exemple permettra ensuite de mieux apprécier les difficultés et les résultats que l'on peut espérer d'un tel protocole.

1. Principe du protocole

Pour Peirce, l'abduction désigne « une certaine capacité de l'esprit humain à deviner l'hypothèse qu'il faut soumettre à l'expérience, laissant de côté sans les examiner la vaste majorité des hypothèses possibles » (Peirce, 1987). Pour des domaines spécialisés, nous poserons que l'expertise est essentiellement une capacité d'abduction. On cherchera alors à approcher cette capacité d'abduction à l'aide de méthodes d'apprentissage connexionnistes. Il suffira enfin de remarquer qu'une telle capacité d'abduction est équivalente à un ensemble de clauses et de proposer des méthodes d'extraction de ces clauses.

1.1. Expertise et abduction

Le médecin qui devine un diagnostic dès les premières réponses d'un patient, le maître international aux échecs qui devine quelques stratégies excellentes parmi quelques milliards de stratégies possibles, et plus généralement l'expert possèdent précisément cette capacité d'abduction dont parle Peirce. Ils utilisent des indices de la situation pour proposer quelques hypo-

thèses en faible nombre. *On s'intéresse dans toute la suite à des experts qui savent identifier suffisamment l'espace A des indices et l'ensemble B des hypothèses.* On peut alors représenter l'expertise comme une fonction « abductive », au sens suivant :

Définition : Une fonction *abductive* est une fonction $f : A \rightarrow 2^B$ qui interprète des indices de A pour produire un certain nombre d'hypothèses dans un espace B discret ¹.

Les valeurs d'une fonction abductive sont des sous-ensembles d'hypothèses. Nous supposons que nous disposons d'une base d'exemples suffisamment importante, pour obtenir par apprentissage supervisé une fonction abductive suffisamment proche du savoir-faire de l'expert.

1.2. Apprentissage supervisé et algorithmes PAB

L'apprentissage supervisé se réalise à partir d'un ensemble de couples (x_i, y_i) . Il consiste à trouver une fonction qui minimise l'erreur quadratique sur la base d'exemples tout en donnant de bonnes performances en *généralisation* sur une base de test de même distribution que la base d'exemples. Sans cette contrainte, une stratégie triviale consisterait à élargir l'espace dans lequel on recherche la fonction et on obtiendrait un apprentissage par cœur de la base d'exemples, mais de mauvais résultats en généralisation sur la base de test. Il existe une théorie et des constatations empiriques sur les liens entre l'erreur en généralisation, la taille de la base d'exemples et le nombre de degrés de liberté du réseau connexionniste (Baum et Haussler, 1989).

On montre par ailleurs que l'erreur quadratique en généralisation est la somme de l'erreur quadratique par rapport aux probabilités bayésiennes à un terme constant près (Gish et Miyake, 1990). Minimiser l'erreur quadratique en généralisation revient donc à estimer les probabilités bayésiennes. *L'apprentissage supervisé apparaît ainsi comme une forme approchée de la théorie de la décision*, lorsque les probabilités de Bayes ne sont pas calculables directement. Remarquons que ce résultat n'est lié ni à la forme de la fonction ni à la procédure de minimisation à la condition qu'elle fournisse une ϵ -optimisation. Plusieurs tentatives ont été faites pour comparer la solution bayésienne, lorsqu'elle est connue, avec les résultats obtenus par des réseaux connexion-

1. Notons aussi, à quelques termes près, le rapport entre cette définition et les espaces d'indices et de significations utilisés par B. Victorri (1988). L'ambiguïté y joue le rôle du petit nombre d'hypothèses, et les dynamiques le rôle de f .

nistes de différents types (Kanaya, 1990) : ces comparaisons montrent qu'il est effectivement possible d'approcher de près la solution bayésienne.

Les remarques précédentes nous ont conduit à définir une classe de procédures admissibles pour l'apprentissage supervisé, les procédures PAB (Bochereau, 1991). Une procédure PAB (Probably Almost Bayesian) est une procédure d'apprentissage supervisé possédant les propriétés suivantes :

a) Étant donné une base d'exemples de taille n , elle recherche une structure de fonction (dans un treillis d'espace fonctionnel, dont l'élément maximal Φ_n dépend de n) et une valeur des paramètres de cette structure qui fournisse une solution ε -bayésienne avec une probabilité $1-\eta$.

b) Quand la taille n de la base tend vers l'infini, Φ_n tend vers un espace fonctionnel dense dans l'espace Φ des fonctions recherchées, et ε et η tendent vers 0.

c) La complexité en temps de la procédure est polynomiale en fonction de n .

De telles procédures sont des algorithmes probabilistes, qui présentent une meilleure complexité au prix de l'introduction d'une certaine probabilité de se tromper. Les conditions (b, c) sont semblables aux conditions posées dans les procédures PAC (Probably Almost Correct) proposées par Valiant (1984). L'originalité de la définition réside dans l'introduction d'un treillis d'espaces fonctionnels, qui permet de formaliser la recherche d'une bonne structure de réseau.

Revenons à présent à notre problème de diagnostic. Nous cherchons une fonction abductive qui mime le comportement de l'expert, tel qu'il apparaît dans une base d'exemples. En utilisant une procédure PAB, on obtiendra les probabilités ε -bayésiennes des hypothèses. Une prise en compte des risques de deuxième espèce conduira à la solution ε -bayésienne.

En fait, les différentes sources d'erreur (nombre d'exemples insuffisants, imprécisions de la procédure de recherche, contraintes arbitraires sur l'espace fonctionnel dans lequel se fait la recherche) introduisent une incertitude sur la solution ε -bayésienne. On préférera souvent retenir un doublé, un tiercé dans l'ordre...

Si on désigne par A l'espace des indices et par B l'ensemble des hypothèses, on obtient alors une fonction « abductive », $f : A \rightarrow 2^B$ (cf. 1.1).

1.3. Fonction abductive comme ensemble de clauses

Parmi les diverses façons de représenter une fonction abductive, il en existe une qui est de nature logique (c'est-à-dire s'écrivant comme un ensemble de clauses) :

Proposition d'équivalence fonctionnelle : Une fonction abductive f est fonctionnellement équivalente à un ensemble de clauses.

Démonstration :

Pour chaque hypothèse $y \in B$, on construit le sous-type $g(y)$ de A :

$$g(y) = \{x : A / y \in f(x)\}$$

$g(y)$ est la restriction du type sur x telle que y appartient aux hypothèses admissibles. On obtient, pour chaque y , l'implication :

$$\forall y \in B \ x \in g(y) \Rightarrow \text{Admissible}(y)$$

La fonction logique s'écrit comme l'ensemble de ces implications.

c.q.f.d.

Ceci reste valable quelle que soit la complication des types A ou B. Remarquons que ces implications ouvrent la voie à un mode d'explication, qui exprime le domaine de validité d'une hypothèse autour du point d'observation. Ainsi, toute machine abductive a la possibilité épistémique de représenter et d'expliquer son fonctionnement de façon logique.

Intéressons-nous à présent à un cas particulier important de la proposition d'équivalence fonctionnelle : les indices et les hypothèses appartiennent à un produit cartésien de types finis discrets et peuvent être représentés comme vecteur de bits; l'ensemble des implications est alors un ensemble de clauses en logique des propositions (Bochereau *et al.*, 1990).

Cette proposition d'équivalence introduit un débat sur deux stratégies de représentation des mécanismes abductifs au cours de l'apprentissage : on peut soit apprendre directement les règles (par exemple, les systèmes de classificateurs génétiques (Goldberg, 1989) soit apprendre au travers d'une fonction continue (par exemple, les réseaux connexionnistes multicouches). C'est le débat « continu vs discret ». Le choix entre ces deux stratégies doit comparer, à performance égale du résultat, les temps d'apprentissage augmentés du temps éventuel d'extraction des clauses.

2. Discussion d'un exemple

Le Conseil d'État est l'une des deux principales juridictions en France. Il peut être saisi par un particulier à propos de la décision d'un maire et doit alors rendre un arrêt sur la validité de la décision municipale. Bien que le droit français, contrairement au droit anglo-saxon ou européen, interdit toute référence explicite à un jugement rendu antérieurement, il apparaît que de nombreuses règles implicites interviennent dans la décision d'un juge du Conseil d'État. En effet, ce dernier, s'inspirant d'un ou plusieurs jugements

rendus antérieurement, est souvent tenté de reproduire les raisonnements qui ont conduit à ces jugements et les adapter aux faits en cours.

Chaque arrêt du Conseil d'État est systématiquement résumé en quelques lignes par un juriste. 350 arrêts du Conseil d'État, échelonnés sur une soixantaine d'années, jugeant un excès de pouvoir ou un non excès de pouvoir de la part d'un maire ont ainsi été sélectionnés puis répertoriés selon une grille prédéfinie (Bourcier, 1988). La codification d'un arrêt fait intervenir plusieurs catégories de classement : (i) la police municipale, (ii) les standards normatifs, (iii) les standards circonstanciels, (iv) la décision du maire et (v) la décision du juge.

2.1. Résultats

On considère le treillis des sous-domaines du domaine de validité. Les domaines d'annulation, de validation et d'hésitation correspondent respectivement à des réponses du réseau connexionniste comprises dans $[0.0, 0.3]$, $[0.3, 0.7]$, $[0.7, 1.0]$. Ces domaines s'expriment à l'aide de clauses, qui sont des éléments maximaux du treillis des sous-domaines homogènes pour la décision (annulation, validation, hésitation).

Tableau 1. — Taille des domaines et nombre de clauses correspondantes.

Domaine	Taille du domaine	Nombre de clauses
Annulation = $[0.0, 0.3]$	22 717 (45 %)	884 (30 %)
Validation = $[0.7, 1.0]$	22 469 (45 %)	809 (25 %)
? = $[0.3, 0.7]$	4 734 (10 %)	1 295 (55 %)
Total	49 920 (100 %)	2 988 (100 %)

On peut remarquer que la taille des domaines de validation et d'annulation sont équivalentes (en nombre de cas) et beaucoup plus importantes que celle du domaine d'hésitation. Inversement, le nombre de clauses est plus grand pour le domaine d'hésitation. Cette inversion est facile à interpréter : il faut un beaucoup plus grand nombre de clauses pour définir une frontière.

Pour étudier de façon plus fine la microstructure de l'expertise, on cherche ensuite à ordonner ces clauses de façon à couvrir la plus grande part des domaines avec un minimum de clauses. C'est une tâche très combinatoire parce que les clauses sont des éléments maximaux qui se recouvrent plus ou moins partiellement. On a choisi d'appliquer un algorithme glouton à la liste des clauses classées par ordre de taille : à chaque itération, on choisit dans la liste la clause qui apporte la contribution complémentaire la plus grande

aux clauses déjà sélectionnées; la complexité de cet algorithme glouton est en $O(n^2)$, si n désigne le nombre de clauses : bien que cette complexité soit raisonnable, il n'a pas été possible de classer plus de 270 clauses avec cet algorithme; et la colonne relative au domaine d'hésitation s'en trouve incomplète sur le tableau suivant.

Tableau 2. — Microstructure de l'ensemble des clauses d'un domaine.

% du domaine	Annulation	Validation	?
10		1	14
20	1	2	39
30	3	3	83
40	5	4	146
50	7	6	270
60	14	11	—
70	24	22	—
80	50	43	—
90	121	109	—
100	884	809	1 295

Le tableau 2 montre, pour chaque domaine, que les nombres de clauses sont géométriquement croissants pour des accroissements linéaires de la taille du domaine. Cela conduit à l'hypothèse suivante :

Hypothèse : le nombre de clauses nécessaire pour représenter une expertise croît exponentiellement avec la précision de cette expertise.

Dans le cas où on peut accepter une hypothèse d'équiprobabilité des cas au sein du domaine de validité, il faut aussi un nombre exponentiellement croissant de clauses pour obtenir une performance linéairement croissante en direction de l'estimateur bayésien (cf. tableau 3).

Tableau 3. — Performance en fonction du nombre de clauses.

Nombre de clauses	Performance
3	55 %
13	62 %
93	70 %
2 988	75 %

2.2. Discussion

La discussion porte d'une part sur le protocole de constitution de la base d'exemples et, d'autre part, sur le protocole d'apprentissage et d'extraction de règles.

La structure du réseau (réseau connexionniste multicouches), le mode d'apprentissage (rétropropagation) et le mode d'extraction retenu (généralisation à partir du calcul exhaustif de tous les cas, comme dans l'apprentissage symbolique) ne sont pas les mieux adaptés dans le but d'une extraction des clauses équivalentes. Pour l'extraction, les méthodes de programmation avec contraintes sont beaucoup plus adéquates du point de vue du temps moyen du calcul. Pour l'apprentissage, les réseaux à membranes (Deffuant, 1991) ont une structure particulièrement adaptée pour l'extraction : cette structure s'interprète comme une union de polyèdres convexes. Leur structure présente d'excellentes propriétés d'évolution grâce à un double mécanisme de recrutement et d'élimination des unités du réseau. Ce double mécanisme prend en compte localement les contraintes de généralisation, telles qu'elles ont été posées par Baum (1989).

Quelles que soient les méthodes d'extraction utilisées, du moment que l'algorithme d'apprentissage est PAB et que l'algorithme d'extraction produit les clauses maximales, les ensembles de clauses obtenus seront voisins (au moins fonctionnellement) et leurs microstructures semblables. Mais il faut encore que la base d'exemples suffise à produire de bonnes généralisations.

En effet, les questions essentielles se situent ailleurs : *a)* dans les signes pris en compte dans le mécanisme d'abduction et *b)* dans la taille et la distribution de la base d'exemples.

Dans le cas du corpus de jurisprudence du Conseil d'État, seul un petit nombre de paramètres multimodaux ont été pris en considération : en conséquence, l'estimation quasi-bayésienne obtenue est médiocre (de l'ordre de 75 %). Une codification plus précise du contenu des arrêts du Conseil d'État serait nécessaire.

Par ailleurs, il n'est pas possible dans ce cas de contrôler la distribution des exemples, en particulier le long des frontières. Mais il serait essentiel de coder l'ensemble des exemples disponibles : en s'appuyant sur une règle empirique proposée par Baum (1989), la précision par rapport au bayésien croît de façon inversement proportionnelle avec le nombre d'exemples.

Conclusion

Dans le cas d'un domaine d'expertise peu formalisé, le protocole d'élicitation des clauses équivalentes décrit par cet article ouvre une des voies possibles pour l'explicitation d'une expertise. Il présente un triple avantage :

a) Le recueil de l'expertise se fait à partir de cas traités par l'expert et non à partir des règles directement explicitées par lui. Or, il semble que la précision de l'expertise croît linéairement avec le nombre d'exemples, en utilisant un apprentissage (Baum, 1989) et logarithmiquement avec le nombre de règles. Si ces hypothèses s'avéraient fondées, elles impliqueraient une préférence de recueil de l'expertise à partir de cas.

b) Si l'algorithme d'apprentissage supervisé produit un réseau quasi-bayésien, les clauses extraites, dans le domaine de validité de l'expertise, se comportent presque toujours comme l'expert. L'algorithme d'extraction des clauses produit une théorie du domaine, directement intelligible par l'expert : ce dernier, en observant les principales règles, peut constater les erreurs principales de cette théorie et proposer de nouveaux indices et de nouveaux exemples susceptibles de les corriger. Dans ce va-et-vient, la motivation de l'expert est d'obtenir sa propre théorie dans son domaine et le processus s'arrête lorsque la fréquence des erreurs par rapport à son propre fonctionnement devient suffisamment faible. On obtient un système hybride, qui apprend de façon connexionniste des exemples et produit des connaissances symboliques (Konfé, 1991).

c) La possibilité effective de construire de grandes bases de clauses permet d'étudier la structure de l'expertise et, en premier lieu, sa répartition entre règles grossières et règles fines. On pourrait les comparer aux règles obtenues à l'aide d'interviews d'experts. Il est ainsi possible de soumettre à l'expérience l'hypothèse plausible que les experts communiquent (et enseignent aux novices) en priorité les « grosses » règles. On pourrait aussi utiliser les ensembles ordonnés de règles dans un but pédagogique.

Le protocole proposé consiste à acquérir le savoir-faire de l'expert avant d'explicitier ce savoir-faire en savoir : ce détour nous semble indispensable, au moins dans le cas où un ou plusieurs experts traitent des exemples de façon contradictoire. Si les difficultés d'ordre computationnel restent importantes dans ce protocole, ce sont les protocoles de recueil d'exemples qui continueront de poser les questions les plus délicates. L'ambition était seulement d'ouvrir une voie dans laquelle différents paradigmes et différentes disciplines des sciences de la cognition peuvent coopérer.

Une telle élicitation de l'expertise ne sera jamais complète : il manquera toujours à la base d'exemples d'autres cas connus de l'expert et d'autres

savoirs utiles à son expertise. Plus grave encore, il subsistera nécessairement des zones d'ombre, au niveau même des indices et des hypothèses non explicités. L'expertise est par nature complexe et ne saurait être représentée exactement par des clauses logiques. Et il n'est pas étonnant que le nombre de clauses croissent exponentiellement en fonction de la précision souhaitée.

Remerciements

Les auteurs tiennent à remercier tous ceux qui leur ont apporté leur concours dans le traitement du corpus de jurisprudence municipale du Conseil d'État, et tout particulièrement Danièle Bourcier qui nous a proposé ce corpus, le groupe de travail de l'ENGREF qui a réalisé l'apprentissage connexionniste et celui du DEA IARFA qui a construit les clauses équivalentes.

Bibliographie

- E. B. BAUM, D. HAUSSLER, What Size Net Gives Valid Generalization? *Neural Computation*, 1, p. 151-160, 1989.
- L. BOCHEREAU, *Conception et interprétation de réseaux connexionnistes multicouches*, Thèse de l'Université de Paris-VI, 1991.
- L. BOCHEREAU, P. BOURGINE, Implémentation et extraction de traits sémantiques sur un réseau neuromimétique, *Actes de Neuro-Nîmes 89*, p. 125-143, 1989.
- L. BOCHEREAU, P. BOURGINE, Extraction of semantic features and logical rules from a multilayer neural network, *Actes de IJCNN90*, Washington DC, 15-19/01/90, 1990.
- L. BOCHEREAU, P. BOURGINE, Validity domain and extraction of rules on a multilayer neural network, *Actes de IJCNN 90*, San Diego, 17-21/06/90, 1990.
- L. BOCHEREAU, P. BOURGINE, H. PRISO, Generalist vs Specialist Neural Networks, *Cognitiva 90*, Madrid, 20-23/11/90, 1990.
- L. BOCHEREAU, P. BOURGINE, G. DEFFUANT, Equivalence between Connectionist Classifiers and Logical Classifiers, *Lecture Notes in Physics*, 368, Springer Verlag, p. 351-363, 1990.
- D. BOURCIER, *Analyse des standards de police municipale : approches décisionnelles*, Thèse d'État, Univ. de Paris, 1988.
- G. DEFFUANT, Perceptron membranes: a self-building connectionist network, soumis pour publication à *Neural Networks*, 1991.
- H. L. DREYFUS, 1972, *Intelligence Artificielle, mythes et réalités*, Flammarion, Paris, 1984.
- H. GISH, A probabilistic approach to the understanding and training of neural network classifiers, *IEEE CH2847-2/90*, p. 1361-1364, 1990.
- D. E. GOLDBERG, *Genetic algorithms*, Addison Wesley, 1989.

- H. H. HOLLAND, Adaptive algorithms for discovering and using general patterns in growing knowledge-bases, *Intern. J. Policy Analysis and Inform. Systems*, 4, (3), p. 245-268, 1980.
- F. KANAYA, S. MIYAKE, Bayes statistical and valid generalization of pattern classifying neural networks, In *proc. Cognitiva 90*, Afcet-Cognitiva, p. 13-19, 1990.
- A. KONFÉ, *Contribution au traitement automatique de la polysémie*, Thèse de l'Université de Caen, juin 1991.
- C. S. PEIRCE, *Textes fondamentaux de sémiotique*, Méridiens Klincksiek, Paris, 1987.
- H. A. SIMON, Problem formulation and alternative generation in the decision making process. *Technical Report AIP 43*, Carnegie Mellon University, Pittsburgh, 1988.
- L. G. VALIANT, A theory of the learnable, *Communications of the ACM*, V27, n° 11, p. 1142-1184, 1984.
- B. VICTORRI, Modéliser la polysémie, *TA Informations*, 29, n° 1-2, p. 21-42, 1988.