

DIANA: système d'aide au diagnostic en culture d'ananas.

X. PERRIER et J.J. LACOEUILHE

Le développement des méthodes informatiques dites d'Intelligence Artificielle et en particulier l'approche Système Expert, permet d'envisager des applications en agronomie dans des domaines qui jusqu'à présent étaient peu accessibles par les méthodes classiques. Le champ d'application qui paraît le plus évident est celui du diagnostic des maladies (l'exemple le plus ancien est TOM développé par l'INRA sur la tomate) ou plus généralement des anomalies observées au cours de la culture.

LE CADRE DE DIANA

L'objet est le diagnostic, en culture d'ananas, de tout désordre d'origine culturale, parasitaire, nutritionnelle, etc.

Les observations sont macroscopiques, réalisables à l'oeil nu, au niveau de la parcelle (pas d'observations microscopiques, d'analyses chimiques, etc., qui peuvent cependant être conseillées, dans les conclusions, pour préciser le diagnostic). Les informations sur l'historique de la parcelle qui sollicitent la mémoire de l'agriculteur, et ne sont pas toujours disponibles, ne sont utilisées qu'en dernier ressort.

Les utilisateurs prévus sont des producteurs ou des techniciens des structures d'encadrement qui ont une bonne connaissance de la plante sans être des spécialistes de chaque discipline.

Le système doit fonctionner sur micro-ordinateur.

POURQUOI UNE APPROCHE SYSTEME-EXPERT ?

Les méthodes classiques de clés-diagnostic produisent des arbres de cheminement où les noeuds sont des questions posées à l'utilisateur et les branches, les différentes réponses possibles. Les questions doivent avoir un nombre limité de réponses, clairement définies (quantitatives ou qualitatives) et immuables. Il est donc tout à fait impossible de gérer des données floues, à réponses ambiguës, incomplètes ou inconnues qui sont souvent les plus fréquentes dans les problèmes de diagnostic agronomique.

SI :

coloration du feuillage = jaune ou rouge
mais pas jaune d'or

et croissance ralentie = oui

et flétrissement = oui

et hétérogénéité = en plage

et plages d'hétérogénéité = plutôt en bordure
de parcelle

ALORS :

symptômes de Wilt - vraisemblance 50

L'approche système-expert (SE) bâtie sur un enchaînement de règles (fait A et fait B entraînent conclusion 1 avec une certaine vraisemblance), permet de prendre en compte ce type d'information.

On pourrait, en schématisant, parler de faisceaux de présomption pour un SE alors qu'une clé-diagnostic exige des preuves.

L'ananas est une plante à parenté épiphyte, son système racinaire est réduit et fragile. Tous problèmes affectant le développement racinaire (parasitisme, particularité pédologique, nutrition) se traduisent par des symptômes comparables correspondant à l'absence d'alimentation de la plante. L'identification des causes réelles à partir de ces symptômes est difficile, c'est souvent un ensemble de facteurs peu informatifs par eux-mêmes qui permettent l'expertise. L'approche SE semble la plus adaptée pour traiter ce type d'information.

UN CONTEXTE FAVORABLE

1. On vient de souligner la difficulté du diagnostic mais des experts se sont intéressés à ce problème et ont essayé de le décomposer. On part donc d'une analyse qui est assez proche, du moins dans le raisonnement, de la structuration de l'information demandée par un SE.

2. L'ananas est une plante qui se prête assez bien à un diagnostic au niveau de la parcelle. La taille est suffisamment petite pour que l'on puisse effectivement observer une population dans son ensemble mais suffisamment grande pour que l'on puisse choisir et étudier en détail un individu.

3. Une variété, Cayenne lisse, domine très largement en particulier dans les zones de culture intensive.

4. Un tel système ne pourra être utilisé directement que par un nombre réduit d'agriculteurs : cependant, il existe souvent dans les pays producteurs des structures professionnelles et des techniciens chargés de l'encadrement, ces techniciens semblent être les utilisateurs désignés de ce système.

STRUCTURATION DE L'INFORMATION

L'organisation générale de la base des connaissances (définition des paramètres et de leurs attributs, contenu des règles et leurs attributs, contraintes sur l'enchaînement des règles) nécessite une structuration préalable de l'information. Cette formalisation n'est pas dictée par le moteur d'inférence (le programme qui, par combinaison judicieuse des règles, simule le raisonnement) mais par les caractéristiques du problème à traiter ; le schéma de raisonnement adopté doit traduire le fonctionnement intellectuel de l'expert.

Deux idées de base ont prévalu lors de la réalisation de DIANA :

1) les anomalies rencontrées sur ananas sont tout autant le fait de l'environnement pédo-climatique ou technique que d'attaques de parasites ou de prédateurs ; l'approche doit être globale. L'expert principal - le chef d'orchestre - est un agronome, les experts spécialisés dans une discipline (physiologie, nématologie ...) n'interviennent qu'en deuxième plan pour étoffer la trame fixée par l'agronome.

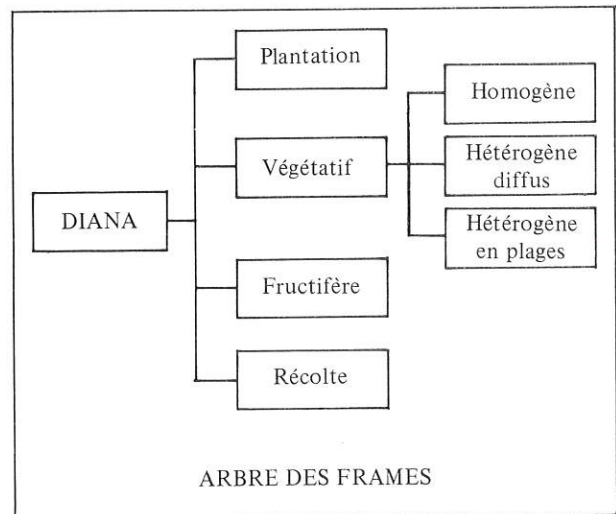
2) il a été souligné le problème des symptômes sur plantes peu informatifs, ils doivent être complétés par des informations issues de l'observation de la population dans son ensemble. L'acquisition de l'information est donc hiérarchisée suivant trois niveaux :

- la parcelle, type d'anomalies, répartition, liens avec les facteurs d'environnement ...,
- la plante, observation de quelques plantes représentatives,
- l'organe, détails sur les parties atteintes.

REALISATION DU SYSTEME

Le logiciel utilisé est PCPLUS de Texas Instruments sous MS DOS, il intègre un moteur d'inférence d'ordre 0+ et un environnement de travail pour le développement. En version utilisateur, il ne nécessite pas d'extension mémoire et fonctionne sur les micro-ordinateurs de type IBM, à partir du 286.

Il autorise le découpage de la base en chapitres homogènes et indépendants (les frames) ; DIANA est organisé en une hiérarchie de frames à deux niveaux. Le premier



distingue les différentes phases de la culture où les problèmes ne sont pas de même nature (plantation, phase végétative, phase fructifère, récolte). Le deuxième niveau concerne le type de répartition des anomalies dans la parcelle (homogène, hétérogène diffus, hétérogène en plages), ces répartitions traduisant des anomalies très différentes.

Actuellement, seules les deux premières phases ont été traitées, la base comprend 120 règles traitant de 62 paramètres. Les anomalies diagnostiquées (les buts) sont au nombre de 29 mais elles débouchent sur 70 messages de conclusion.

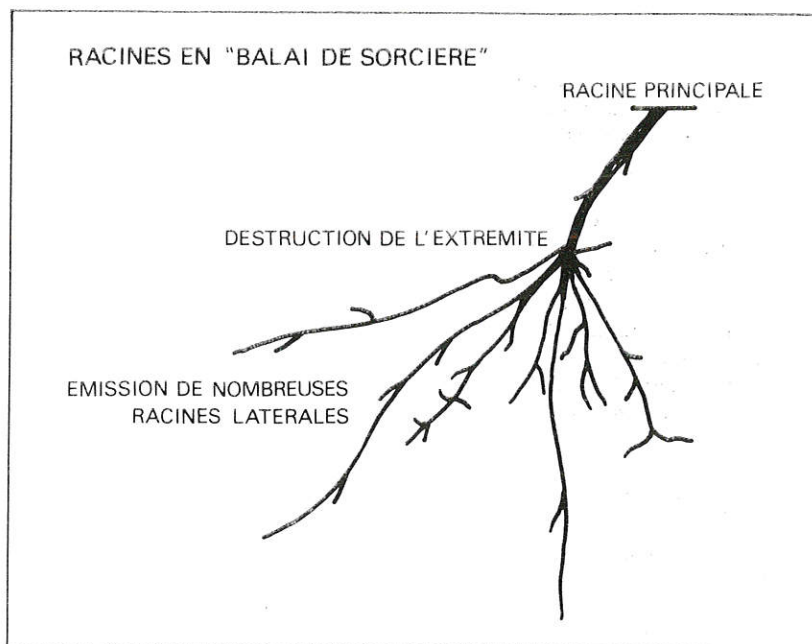
En effet l'objet n'est pas simplement d'identifier une anomalie mais d'en déterminer si possible l'origine. Par exemple, une coloration vert foncé du feuillage est, dans un certain contexte, caractéristique d'une nutrition azotée excessive ; son origine peut être une fertilisation azotée excessive mais également un déséquilibre nutritionnel, en particulier avec le phosphore. On cherchera donc des informations sur les fertilisations pratiquées ou les précédents culturels avant de conclure sur l'anomalie et sa cause.

INTERFACE UTILISATEUR

L'utilisateur visé n'est pas un spécialiste des différents problèmes. Cela impose un effort particulier pour donner au système un abord simple et convivial incitant à la persévérance.

La pertinence des réponses dépend en grande partie de la formulation des questions posées, elles doivent être simples, non ambiguës et clairement documentées. PCPLUS permet d'associer aux différents paramètres des images graphiques qui sont souvent plus explicites qu'un long texte.

Les conclusions sont le plus souvent agrémentées de commentaires sur le problème détecté afin de donner un rôle pédagogique à ce SE. A la limite, un utilisateur ayant utilisé longuement le système devrait avoir acquis les connaissances nécessaires et le mode de raisonnement pour ne plus avoir besoin de l'outil informatique.



— D I A N A I — IRFA-CIRAD

Conclusion possible à 85 % :

Attaque de champignon (*Phytophthora nicotiana* var. *parasitica*) au cœur du plant, favorisée lorsqu'il y a stagnation d'eau libre dans le sol (mauvais drainage) ; la plantation sur polyéthylène est également un facteur de risque. Les sols à pH élevé (supérieur à 5 ou 5,5) sont très favorables à ce champignon. Les couronnes sont particulièrement sensibles.

Le souci pédagogique a conduit à introduire des questions inutiles pour le diagnostic mais qui permettent d'attirer l'attention sur certains points. Par exemple, lorsque des colonies importantes de cochenilles ont été notées, il n'est pas utile de savoir si on observe des populations élevées de fourmis, ce détail a pourtant été conservé pour attirer l'attention sur le mode d'infestation de ces cochenilles et sur la nécessité de traiter ces populations de fourmis.

LES DIFFICULTES DE REALISATION

Le logiciel PCPLUS offre une interface de développement relativement simple et efficace ; il est peut-être limité dans ses fonctionnalités mais l'ensemble des propriétés que l'on peut attribuer aux règles et aux paramètres permet le plus souvent de contourner les difficultés rencontrées. De plus, les systèmes de type diagnostic sont sur le plan informatique assez simples : pas de saisie en temps réel, pas d'accès à des bases de données ...

Les principaux problèmes ne sont pas, pour cette application, de type informatique mais résident beaucoup

plus dans la collecte et la structuration de l'information (travail du «cogniticien») :

1. récupérer l'information auprès des experts en essayant de suivre un fil logique mais sans imposer de contraintes stérilisantes ;
2. l'organiser en sachant que le niveau de connaissance peut être très variable suivant les domaines. Un intérêt annexe de ce style d'exercice est déjà de mettre en évidence les points où des développements seraient nécessaires ;
3. traiter ce qui apparaît comme des contradictions entre experts et qui génère souvent une information complémentaire pertinente.

Sur le plan pratique, de nombreux points sont à surveiller, en particulier :

1. conserver une apparence logique à l'enchaînement des questions pour éviter de désorienter l'utilisateur ;
2. limiter le nombre de questions au maximum à 7 ou 8 pour arriver à une première conclusion, quitte à affiner par la suite le diagnostic ;
3. maîtriser l'évolution des coefficients de vraisemblance qui sont automatiquement recalculés au fil des questions et peuvent atteindre des valeurs peu réalistes ;
4. éviter les branches qui ne se terminent que par une non conclusion, elles peuvent devenir abondantes lorsque le nombre de paramètres augmente.

DEVELOPPEMENTS PREVUS

On peut considérer que les deux tiers du système sont réalisés, il faut maintenant développer les parties concer-

nant la floraison et la récolte.

L'étape suivante sera la validation en pratique du système, d'une part pour tester son efficacité à poser de bons diagnostics et ensuite s'assurer de son acceptation par les utilisateurs.

L'enrichissement progressif du système sera indispensable d'une part pour y intégrer les nouvelles connaissances mais également pour élargir son domaine d'application. Actuellement il s'adresse principalement à des agriculteurs possédant déjà un certain savoir-faire, pratiquant une culture relativement intensive et situés dans la zone Afrique équatoriale ou Caraïbes. Il devra, dans l'avenir, intégrer d'autres zones écologiques et d'autres systèmes de production.

L'enrichissement de la base, ou sa simple maintenance, pose en pratique, de gros problèmes. Contrairement aux présentations simplistes, une base de connaissance n'est pas un simple empilement de règles indépendantes ; dès que l'on modifie une information, c'est un ensemble de règles qu'il faut modifier et réorganiser. Une structuration très rigoureuse de la base dès le départ peut limiter les déboires liés à ces modifications mais ne les supprimera pas entièrement. Ceci apparaît comme l'un des facteurs limitants le plus important de l'approche SE en diagnostic agronomique; il peut conduire à des systèmes figés, n'évoluant plus et perdant rapidement tout intérêt.

A terme, il serait souhaitable de compléter les conclusions par des conseils d'intervention ; ce pourrait être un outil efficace de passage de l'innovation technique. Cependant des conseils efficaces doivent prendre en compte le contexte économique et social de l'agriculteur et perdent

donc toute généralité. Il sera sans doute nécessaire de créer plusieurs fichiers de conseils, adaptés aux grands types de systèmes de production.

CONCLUSIONS

La difficulté du diagnostic en culture d'ananas demande de la part de l'expert une démarche intellectuelle appropriée. Cette démarche qui a parfois des aspects intuitifs, subjectifs, empiriques, se prête mal à une formulation sous forme de note, guide, mémento. Au stade actuel de développement de DIANA, il semble qu'une approche de type système-expert permette de simuler ce raisonnement avec une certaine efficacité, d'où la possibilité d'en faire un outil au service des producteurs.

C'est également un moyen d'assurer la transmission du savoir d'experts qui, trop souvent, changent d'activité sans avoir pu réellement faire passer leurs connaissances.

Sur le plan de la vulgarisation et du développement, ce type de système présente un intérêt certain : par son rôle pédagogique, il peut améliorer la formation des utilisateurs ou de leurs conseillers ou même servir de support à un enseignement. S'il est régulièrement enrichi et alimenté de nouvelles connaissances, il peut constituer un atout supplémentaire pour la diffusion des innovations techniques.

Sur le plan recherche, la collecte et la structuration préalable de l'information mettent rapidement en évidence les domaines moins bien connus ou étudiés uniquement dans certaines conditions. Ce peut être une aide efficace pour évaluer les priorités des thèmes à développer.