

Le blé dur à l'ère du numérique: Apports de l'ingénierie des connaissances

Rallou Thomopoulos, Patrice Buche

Pour relever le défi d'une filière durable, garante de la viabilité de la situation des acteurs et de la sécurité alimentaire des consommateurs, il importe d'adopter une approche concertée et intégrative entre l'amont et l'aval de la filière. Un premier verrou de cette intégration est la mise en place de systèmes d'information permettant de consulter et d'échanger les données obtenues entre acteurs et champs disciplinaires différents sur l'ensemble de la filière. Un second verrou est la mise en évidence de la diversité des objectifs et des points de vue constituant la filière et leur prise en compte concertée. La levée de ces verrous est un passage nécessaire à la construction d'une vision systémique, pluridisciplinaire et durable de la filière.

Des méthodes et des outils issus de l'ingénierie des connaissances ont montré leur pertinence pour lever plusieurs de ces verrous. Dans ce chapitre, nous présenterons des exemples de résultats de travaux innovants dans l'étape d'intégration des connaissances :

- Utilisation combinée de données expérimentales et de dires d'experts pour comprendre l'impact des opérations unitaires sur la qualité des produits de la filière.
- Identification d'informations subjectives dans le processus d'acquisition de connaissances.

Nous présenterons ensuite plusieurs exemples de méthodes et d'outils originaux facilitant l'exploitation des connaissances pour aider à la prise de décision :

- Simulation de scénarii de type « What-If » (Que se passe-t-il si...?) combinant argumentation et simulation numérique.
- Méthode d'explication de connaissances implicites ou incomplètes

L'enjeu des travaux menés est une meilleure maîtrise de la filière dans le but de pouvoir garantir la qualité, la sécurité, l'accessibilité, la durabilité des produits. La filière blé dur apparaît relativement bien connue et explorée via différentes disciplines complémentaires, si bien qu'une approche intégrative semble envisageable. En effet, les travaux menés ont conduit à l'accumulation de données issues de sources et de disciplines variées (technologie, nutrition, science des aliments) ; c'est la capacité à intégrer ces connaissances qui apparaît comme un objectif prioritaire pour permettre l'aide à la décision.

I. Comprendre l'impact des opérations unitaires sur la qualité

I.1. Intégration de sources de données hétérogènes afin de préparer les données pour l'aide à la décision

Dans cette étude (Thomopoulos et al., 2009 ; Thomopoulos et al., 2010), deux types d'information sont représentés :

1. des dires d'experts, connaissances à caractère générique, découlant de l'expérience des spécialistes du domaine et décrivant les mécanismes communément admis régissant ce domaine. Ces connaissances sont représentées sous la forme de règles dans un modèle graphique et logique, les graphes conceptuels ;
 2. des données expérimentales, issues de la littérature internationale du domaine. Elles sont représentées dans le modèle relationnel. Ces données nombreuses décrivent avec précision et de façon chiffrée des expériences réalisées pour approfondir la connaissance du domaine et leurs résultats. Ces résultats peuvent être en accord ou non avec les connaissances apportées par les direx d'experts.
- La coopération des deux types de connaissances permet de consolider l'expertise du domaine.

Les données expérimentales sont regroupées au sein d'une base de données. La consultation et la saisie des données par des spécialistes du domaine se fait via un navigateur, au travers de formulaires. La base de données contient à l'heure actuelle une trentaine de tables et les résultats détaillés d'environ 600 expériences. Les règles expertes sont représentées à l'aide d'une interface graphique (<http://www.lirmm.fr/cogui/>). Environ 150 règles expertes sont disponibles, une vingtaine est utilisée pour tester la méthodologie proposée. Un protocole de communication entre les deux systèmes est établi.

Les données expérimentales. Destiné à des scientifiques et à des industriels de l'agroalimentaire, l'outil (en langue anglaise) de consultation et de saisie des données expérimentales intègre des données scientifiques, aussi exhaustives que possible, issues de la littérature traitant des qualités nutritionnelles des aliments à base de blé dur, et décrivant l'impact des procédés de transformation sur ces qualités. Un tel article scientifique comporte généralement les informations suivantes :

- des mesures expérimentales, notamment sur l'analyse des nutriments (par exemple : dosage de la vitamine B1 dans les pâtes) ;
- des résultats sur l'impact d'une ou plusieurs opérations unitaires sur une ou plusieurs qualités nutritionnelles (effet de la cuisson-extrusion sur la teneur en minéraux) ;
- des données sur l'influence de certains paramètres de l'opération unitaire (par exemple : l'effet de la température de stockage sur la rétention des vitamines) ;
- des données sur l'influence d'autres opérations unitaires (par exemple : l'effet du type de séchage des pâtes sur la rétention des vitamines pendant la cuisson dans l'eau) ;
- des modèles décrivant l'évolution des qualités nutritionnelles (par exemple : la cinétique de dégradation thermique de la vitamine B1) ;
- des références bibliographiques.

Les connaissances expertes. Les connaissances expertes, représentées par des règles de graphes conceptuels, expriment l'impact connu de chaque opération unitaire sur chaque nutriment, en explicitant les conditions dans lesquelles cet impact se produit (paramètres de l'opération ou interaction avec d'autres opérations unitaires). L'impact peut concerner la variation de la teneur du nutriment (augmentation, diminution, stagnation) mais aussi la modification de propriétés du nutriment, telles que la digestibilité, l'allergénicité, etc.

Les figures 1 et 2 illustrent l'utilisation combinée d'une règle experte et des données expérimentales.

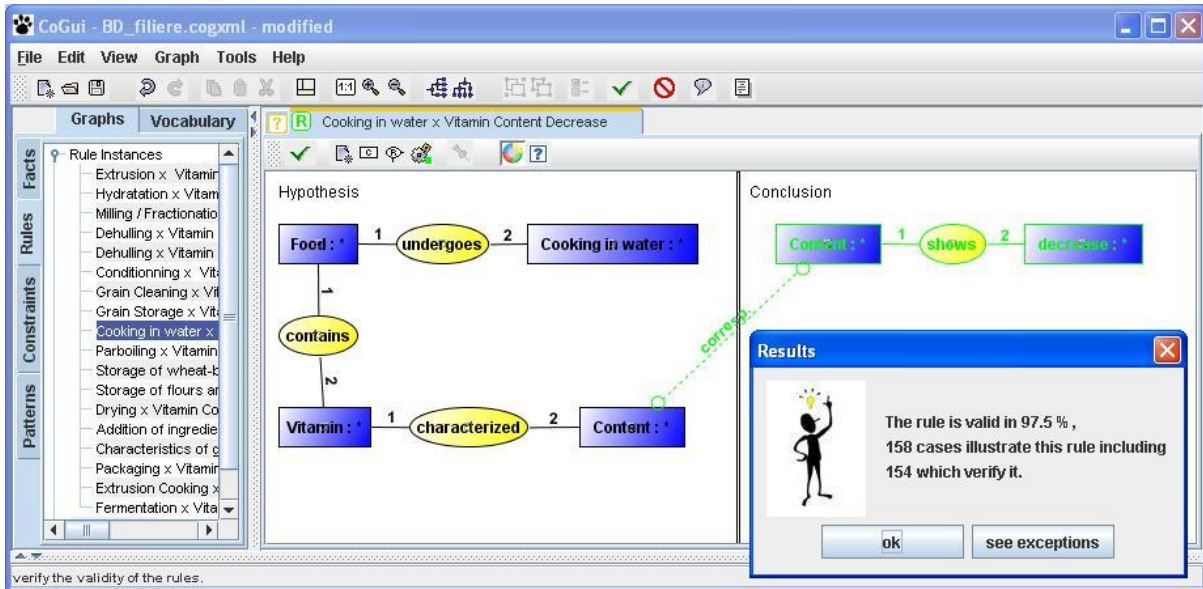


Figure 1 : Recherche des cas particuliers d'une règle experte. La règle est représentée par deux graphes conceptuels (l'hypothèse et la conclusion) et exprime : « si un aliment subit une cuisson à l'eau, alors sa teneur en vitamines diminue ». La fenêtre résultat montre le taux de validité de cette règle, ici 97.5%. Le taux de validité des règles testées varie entre 73% et 100% dans cette application.

Cooking in water x Vitamin Content Decrease
RULE EXCEPTIONS

Food	Parameters of the unit operation				Interactions with prior unit operations	Name of the component	Results		Reference	
	Temperature(°C)	% of salt in water(%)	Time(mn)	Kind of water()			Percentage Value	Standard deviation		Kind of result
Pasta	100	Undefined	20	Undefined	<input type="checkbox"/> Yes	Vitamin A	4%		increase	Parrish, D. B. et al.,1980
Pasta	100	Undefined	20	Undefined	<input type="checkbox"/> Yes	Vitamin A	8%		increase	Parrish, D. B. et al.,1980
Pasta	100	Undefined	20	Undefined	<input type="checkbox"/> Yes	Vitamin A	8%		increase	Parrish, D. B. et al.,1980

Figure 2 : Affichage des cas particuliers. La 1^{ère} ligne du tableau, par exemple, montre qu'une pâte cuite à 100°C pendant 20 minutes conduit à un accroissement de 4% en vitamine A.

I.2. Apport d'informations expertes dites « subjectives » dans le processus d'acquisition de connaissances

Une approche collaborative entre les connaissances expertes et l'exploration de données est un bon moyen de construire des modèles interprétables ayant l'adhésion et la confiance des experts. Nous avons proposé une approche interactive et itérative, guidée par les données et s'appuyant sur un vocabulaire structuré par des relations, appelé « ontologie », pour concevoir des modèles pertinents (ici arbres de décision). Cette approche est présentée plus en détail dans (Johnson et al., 2010; Thomopoulos et al., 2013).

Cette méthode de construction de modèles guidés par les données est :

- collaborative, puisqu'elle fait interagir méthodes d'apprentissage et experts ;
- itérative, car elle implique plusieurs cycles pour améliorer les résultats obtenus ;
- hybride, du fait qu'elle s'appuie à la fois sur les données et les connaissances.

L'approche proposée vise trois objectifs connexes : (i) trouver des variables explicatives pertinentes, (ii) structurer et enrichir les connaissances du domaine sous la forme d'une ontologie, (iii) accroître la confiance des experts dans le modèle.

Son principe est le suivant. Partant d'un ensemble de données initial et de connaissances représentées dans une ontologie, un premier modèle guidé par les données est appris (étape 1). Ce modèle est d'abord évalué avec des critères numériques objectifs. Puis il est soumis aux experts du domaine, qui peuvent enrichir l'ontologie en suggérant de nouvelles relations entre certaines variables (étape 2). Des transformations sont appliquées aux données selon ces nouvelles relations (étape 3). L'objectif est de remplacer les variables jugées non pertinentes par les experts par des variables significatives. L'ensemble du processus est répété de façon itérative jusqu'à ce qu'aucune possibilité d'amélioration ne soit plus détectable. Ce processus peut aboutir à un modèle de prédiction précis, conforme à la fois aux connaissances expertes et aux données. La figure 3 décrit ce processus.

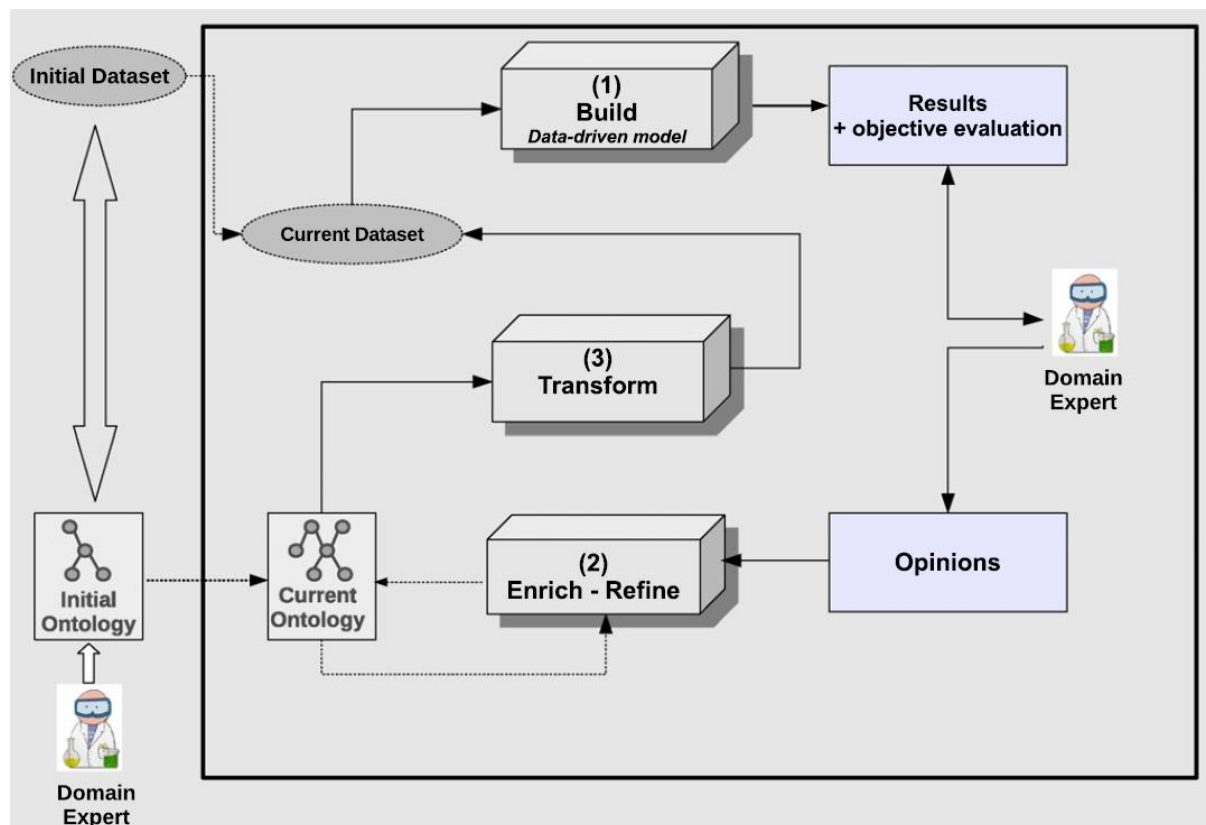


Figure 3 : Schéma du processus de construction de modèle. A partir du jeu de données initial et de l'ontologie initiale construite avec un expert du domaine (à gauche), trois étapes sont appliquées de façon itérative (en boucle dans l'encadré) : 1) apprentissage du modèle à partir des données ; 2) enrichissement et affinement de l'ontologie ; 3) transformation des données selon la nouvelle ontologie.

Les informations concernant l'impact de chaque opération unitaire du procédé de transformation sur chaque type de propriété du produit, se trouvent sous la forme d'un ensemble de données.

Les variables d'entrée sont les paramètres de l'opération unitaire. La variable de sortie est l'impact de l'opération sur une propriété (par exemple la variation de la teneur en vitamines). Nous étudions ici le cas de l'opération unitaire Cuisson à l'eau et de la propriété Teneur en vitamines. Ce cas concerne 150 données expérimentales et implique 60 concepts de l'ontologie. La transformation des données et les arbres de décision ont été obtenus avec le logiciel R (R Core Team, 2016) (utilisation du package R-WEKA et environ 2000 lignes de code développé).

La figure 4 montre l'arbre initial obtenu à partir du jeu de données brutes. La figure 5 montre l'arbre final, où certaines variables initiales mesurées et continues, comme *Cooking time*, sont maintenant remplacées par des variables jugées plus significatives et pertinentes, telles que *Cooking type*.

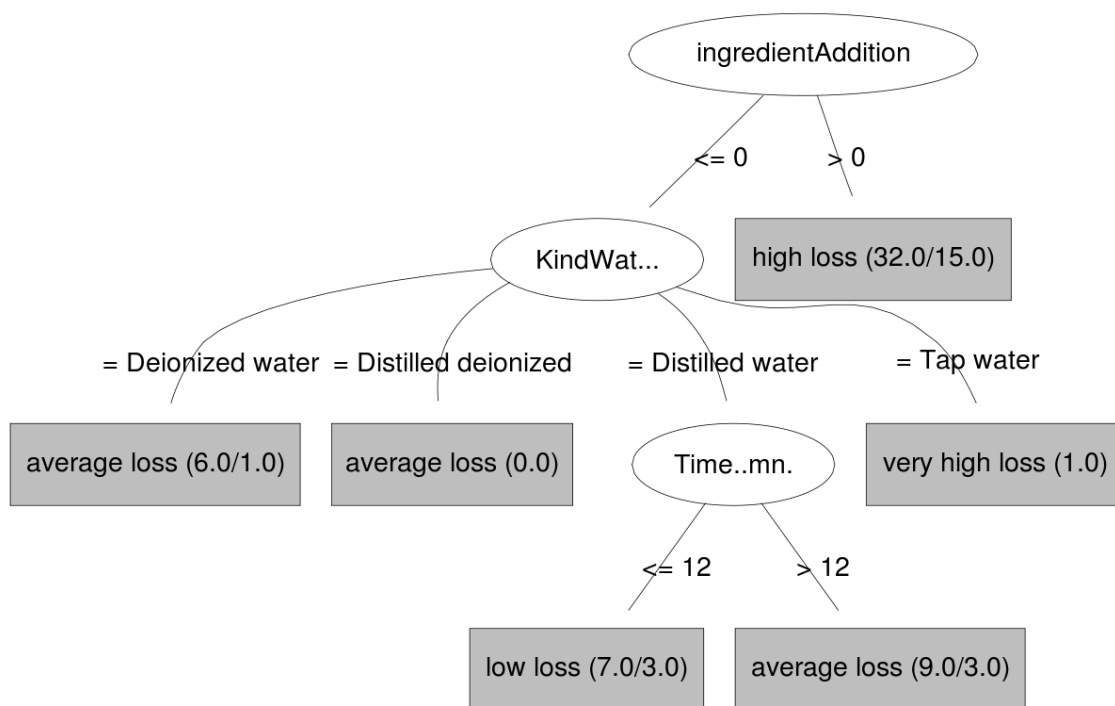


Figure 4 : Arbre de décision à l'état initial

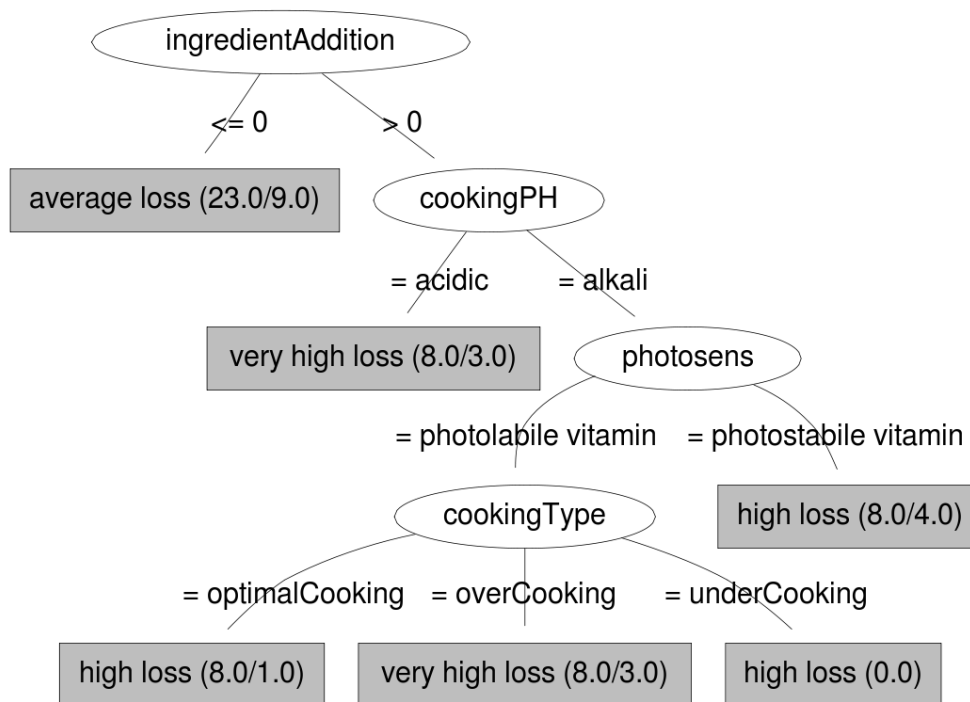


Figure 5 : Arbre de décision à l'état final

II. Exemples de méthodes et d'outils d'aide à la décision

II.1 Simulation de scénarii de type What-If combinant argumentation et simulation numérique

Les filières agroalimentaires impliquent de nombreux acteurs, de la production à la consommation en passant par la transformation, la distribution ou encore le recyclage. En conséquence, outre l'échelle du décideur, les intérêts de chaque partie prenante de la filière interviennent et, compte tenu de la diversité de leurs positions, tendent vers des objectifs souvent non conciliables. Prendre une décision impliquant plusieurs acteurs aux objectifs divergents nécessite de considérer des informations tant qualitatives – les préférences des acteurs sur les décisions possibles – que quantitatives – les paramètres servant d'indicateurs pour les acteurs. Nous nous intéressons à l'association de ces deux types d'approches. Nous simulons ainsi les scénarios découlant de chaque décision par un modèle numérique, pour permettre aux parties prenantes de mieux appréhender les conséquences de certaines décisions (soutenues ou rejetées par les arguments) sur les activités des acteurs.

L'étude concerne les cultures associées de type céréale-légumineuse (Thomopoulos et al., 2018 ; Thomopoulos et al., 2017a ; Thomopoulos et al., 2017b ; Thomopoulos et al., 2016a ; Thomopoulos et al., 2016b) et plus spécialement entre le blé dur et une légumineuse comme le pois, dans le sud de la France. Nous avons établi dans la figure 6 un modèle systémique global de l'organisation du système blé dur-légumineuse et des principaux arguments cités en faveur (couleur rose et symbole '+++') ou en défaveur (couleur verte et symbole '- - -') de la culture associée.

Classiquement, un processus d'argumentation suit trois étapes principales : 1) la construction d'arguments et de contre-arguments, 2) l'évaluation de l'acceptabilité des arguments, 3) la définition de conclusions justifiées. Dans la lignée des travaux existants (Amgoud et Prade, 2009) et pour appréhender les schémas de décision du monde réel, nous adoptons un modèle de décision argumentée dans lequel un argument fournit une raison pour soutenir une décision. Nous y intégrons un ensemble de toutes les options possibles, considérées comme mutuellement exclusives et soutenues par des arguments. Dans le cadre de (Dung, 1995), une relation binaire appelée « attaque » est définie sur l'ensemble des arguments, reflétant les conflits entre arguments. Elle est représentée graphiquement par des flèches entre les arguments en conflit, représentés par des cercles. La figure 7 représente le graphe d'argumentation débattant de l'intérêt de trier ou non les graines produites dans les cultures associées.

Les sémantiques d'acceptabilité de Dung permettent d'identifier, parmi les arguments en conflit, ceux qui seront retenus pour déterminer les décisions acceptables. A partir d'un système de décision argumentée, nous pouvons donc définir un résultat décisionnel, qui fournit en sortie de calcul un ensemble de décisions fondées. Dans la figure 7, la partie de droite montre les arguments les plus précis, discutant spécifiquement de deux options de tri post-récolte, classique ou optique. L'application du principe d'acceptabilité de Dung valide 4 arguments en faveur du tri optique et un seul en faveur du tri classique, pour les cultures associées.

Parmi les arguments validés en faveur du tri optique, il est intéressant de noter que l'un d'eux (argument 15) fait une hypothèse sur l'évolution attendue des prix des trieurs optiques dans les années à venir. Pour aller plus loin, ce type d'argument « futuriste » peut être exploré par la simulation numérique. Dans la suite de la démarche, nous nous sommes donc appuyés sur cet argument pour réaliser des simulations numériques de la marge semi-directe de l'agriculteur en fonction du prix des trieurs optiques, sous des hypothèses optimistes ou pessimistes.

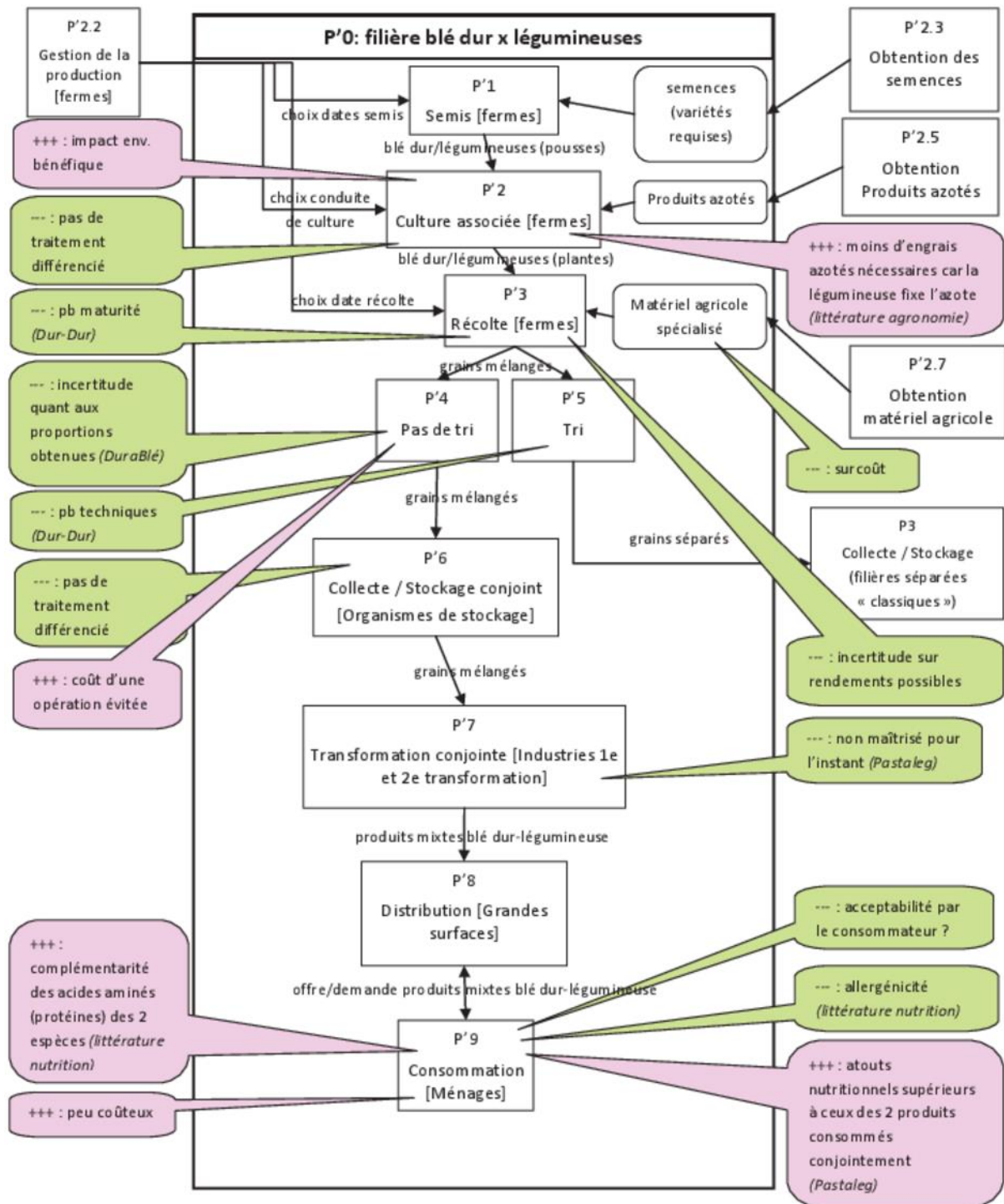


Figure 6 : Schéma du système blé dur-légumineuse avec arguments

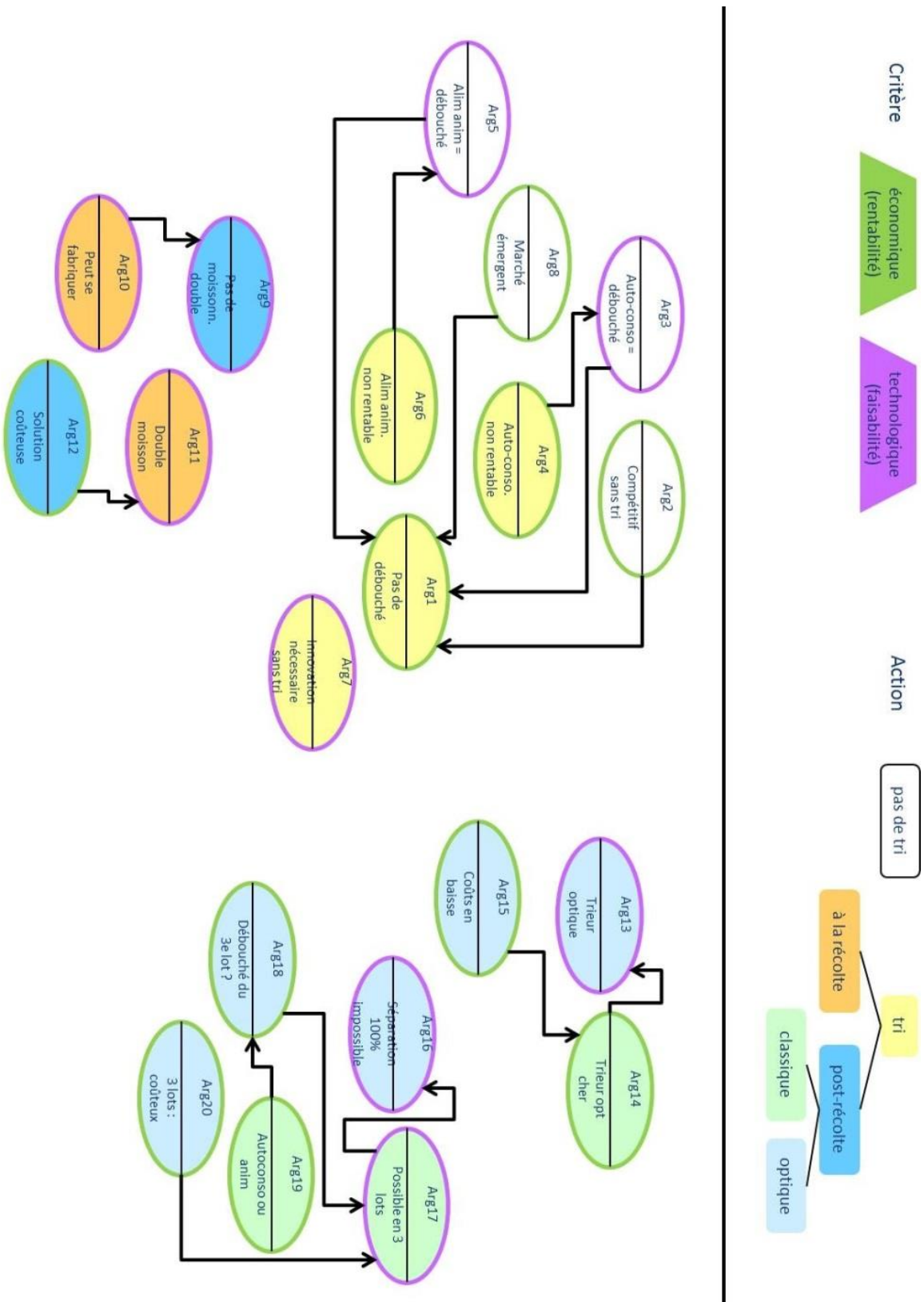


Figure 7 : Arguments économiques et technologiques en faveur ou non du tri des graines dans les cultures associées

II.2 Méthode d'explication de connaissances implicites ou incomplètes

Les filières agroalimentaires impliquent de nombreux acteurs, de la production à la consommation en passant par la transformation, la distribution ou encore le recyclage. Dans le cadre d'une approche systémique, il est nécessaire d'intégrer des connaissances pluridisciplinaires (par exemple agronomiques, sur les procédés de transformation et socio-économique) dans une base de connaissances pour aider à la prise de décision. Cette base de connaissance doit pouvoir être consultable par tous les acteurs, quel que soit leurs connaissances de la filière. Pour faciliter cette consultation, nous proposons un module d'interrogation intégrant des fonctions de dialogue (Arioua et al. 2016a, Arioua et al. 2016b, Arioua et al. 2017). La méthode de dialogue proposée a pour objectif de gérer le fait que certaines des connaissances intégrées dans la base sont implicites ou incomplètes. Dans un mode d'enrichissement de la base, la solution proposée engage l'expert (qui a produit la connaissance) et la machine (qui utilise la base) dans un dialogue argumentatif et explicatif dont l'objectif est de faire expliciter les connaissances implicites. Seul le dialogue explicatif est autorisé dans le mode de consultation simple pour les utilisateurs non experts. Les deux exemples ci-dessous illustrent un dialogue homme-machine lors de la consultation d'un itinéraire technique pour la culture de blé dur dans le cadre du projet ANR Dur-Dur.

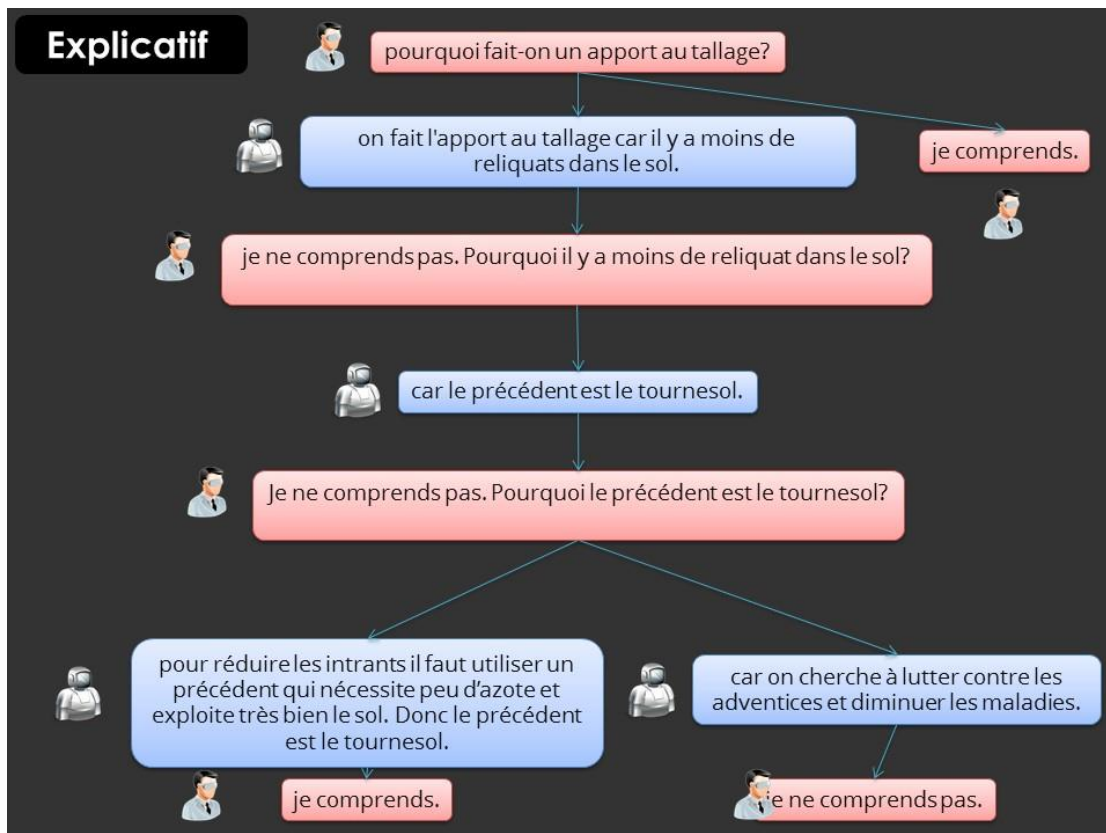


Figure 8: Exemple de dialogue explicatif

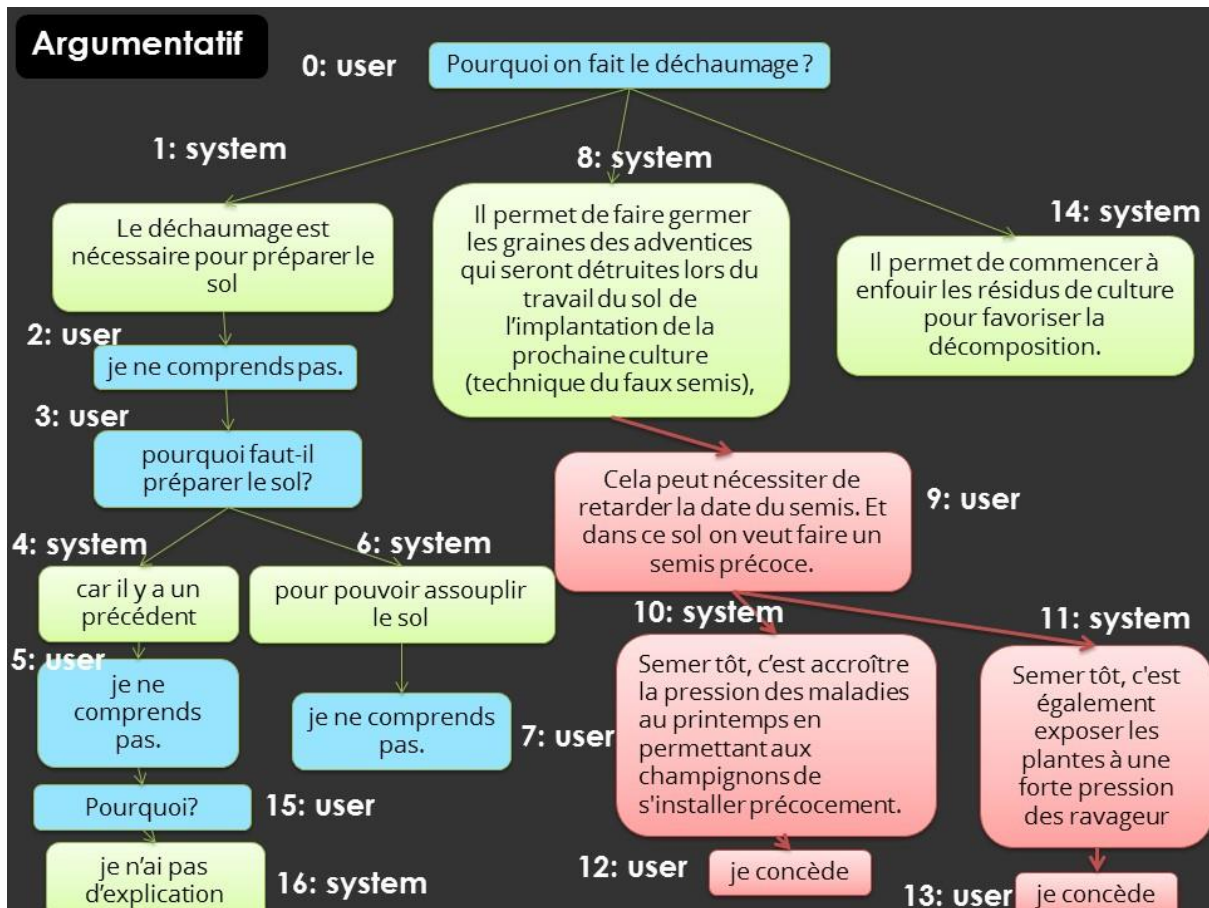


Figure 9: Exemple de dialogue explicatif et argumentatif

Conclusion

Ce chapitre s'est appuyé sur divers exemples issus de la filière blé dur pour illustrer l'intérêt dans ce secteur de méthodes innovantes issues de l'intelligence artificielle, telles que l'utilisation combinée de données expérimentales et de dires d'experts pour maîtriser la qualité des produits de la filière, le rôle des informations subjectives, les possibilités de simulation de scénarii ou encore l'explication de connaissances. Il montre comment l'ensemble des études présentées dans cet ouvrage peuvent être intégrées et fournir matière à une démarche d'aide à la décision.

Références

- R. Thomopoulos, B. Charnomordic, B. Cuq and J. Abécassis (2009). Artificial intelligence-based decision support system to manage quality of durum wheat products. *Quality Assurance and Safety of Crops & Foods* 1(3), 179-190
- R. Thomopoulos, J.R. Bourguet, B. Cuq and A. Ndiaye (2010). Answering questions that may have answers in the future: a case study in food science. *Knowledge Based Systems* 23(5): 491-495
- I. Johnson, J. Abécassis, B. Charnomordic, S. Destercke and R. Thomopoulos (2010). Making Ontology-Based Knowledge and Decision Trees interact: an approach to enrich knowledge and

- increase expert confidence in data-driven models. In Proceedings of KSEM'2010, Belfast, UK, September 2010, *Lecture Notes in Artificial Intelligence* 6291: 304-316, Springer
- R. Thomopoulos, S. Destercke, B. Charnomordic, I. Johnson and J. Abécassis (2013). An iterative approach to build relevant ontology-aware data-driven models. *Information Sciences* 221: 452-472
- R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- L. Amgoud and H. Prade (2009), Using arguments for making and explaining decisions. *Artificial Intelligence* 173(3-4): 413-436
- P.M. Dung (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artificial Intelligence Journal*, 77, 321-357
- A. Arioua, M. Croitoru and P. Buche (2016a) DALEK: A Tool for Dialectical Explanations in Inconsistent Knowledge Bases. COMMA 2016: 461-462
- A. Arioua, P. Buche and M. Croitoru (2016b): A Datalog± Domain-Specific Durum Wheat Knowledge Base. MTSR 2016: 132-143
- A. Arioua, P. Buche and M. Croitoru (2017) Explanatory dialogues with argumentative faculties over inconsistent knowledge bases. *Expert Syst. Appl.* 80: 244-262
- R. Thomopoulos, B. Moulin and L. Bedoussac (2018). Supporting decision for environment-friendly practices in the agri-food sector: when argumentation and system dynamics simulation complete each other. *International Journal of Agricultural and Environmental Information Systems* 9(3): 1-21
- R. Thomopoulos, B. Moulin and L. Bedoussac (2017a). Combined Argumentation and Simulation to Support Decision: Example to Assess the Attractiveness of a Change in Agriculture. In: S. Benferhat et al. (Eds.): IEA/AIE 2017, Part II, LNAI 10351, pp. 275–281, Springer
- R. Thomopoulos, B. Moulin and L. Bedoussac (2017b). Combining argumentation and system dynamics simulation for decision support in the agri-food sector. Proceedings of 2017 EFITA WCCA Congress, Montpellier, July 2017
- R. Thomopoulos, L. Bedoussac and B. Moulin (2016a). Evaluating cereal-legume intercropping towards sole crops by combining argumentation and simulation. In Second International Legume Society Conference, Tróia (Portugal), October 2016
- R. Thomopoulos, S. Gaucel, B. Moulin (2016b). Associer argumentation et simulation en aide à la décision : Illustration en agroalimentaire. Actes de la 16^{ème} conférence internationale sur l'extraction et gestion des connaissances (EGC'2016), Reims, Janvier 2016, *Revue des Nouvelles Technologies de l'Information*, pp. 351-356

Glossaire

Affinement des connaissances : Explicitation de connaissances plus précises que celles initialement disponibles.

Approche itérative : Méthode qui procède par itérations au cours desquelles elle détermine des solutions qui se rapprochent graduellement de la solution cherchée.

Graphes Conceptuels : Modèle de représentation des connaissances et de raisonnement issu de l'intelligence artificielle, de la famille des réseaux sémantiques.

Ontologie : Vocabulaire d'un domaine structuré en concepts et en relations que ces concepts établissent entre eux (spécialisation, règles, etc.).