

Proposition de thèse en représentation des connaissances INRA Montpellier (UMR IATE) / LIRMM (équipe RCR)

Sujet

Raisonnement par priorités et compromis en présence d'informations contradictoires : application à la définition de la qualité alimentaire.

Résumé

Le travail proposé concerne la représentation de connaissances, le raisonnement et la prise de décision en présence de différents points de vue émanant d'acteurs dont la perception et les priorités sont divergentes.

L'objectif du travail est la conception d'un cadre théorique et méthodologique permettant l'aide à la décision au sein d'un modèle de représentation de connaissances.

Le domaine d'application plus particulièrement considéré est la définition de la qualité alimentaire dans les filières agroalimentaires, pour laquelle entrent en jeu différents points de vue (par exemple l'intérêt nutritionnel, les qualités gustatives, la sécurité sanitaire des produits) et différents acteurs (industriels, chercheurs, consommateurs) dont les objectifs divergent.

Contacts

En représentation des connaissances (encadrants de la thèse)

- M.L. Mugnier, LIRMM - RCR, mugnier@lirmm.fr
- R. Thomopoulos, INRA - UMR IATE / LIRMM - RCR, rallou@supagro.inra.fr

Concernant le domaine d'application

- B. Cuq, Montpellier SupAgro - UMR IATE, cuq@supagro.inra.fr
- J. Abécassis, INRA - UMR IATE, abecassi@supagro.inra.fr

Description détaillée

Voir page suivante

Financement

½ bourse INRA (obtenu) et ½ bourse Région (demandé)

Bibliographie

Marie-Laure Mugnier, Michel Chein. *Représenter des connaissances et raisonner avec des graphes*, Revue d'Intelligence Artificielle, vol. 10(1), pages 7-56, 1996.

Denis Bouyssou, Didier Dubois, Marc Pirlot, Henri Prade. *Concepts et méthodes pour l'aide à la décision*, Lavoisier, 3 volumes, Traité IC2, 2006.

Rallou Thomopoulos, Patrice Buche, Ollivier Haemmerlé. *Fuzzy Sets Defined on a Hierarchical Domain*, IEEE Transactions on Knowledge and Data Engineering, 18(10), pp. 1397-1410, 2006.

Rallou Thomopoulos, Bernard Cuq et Joël Abécassis. *Integrating Experimental Data and Expert Knowledge to Capture the Impact of Food Processing on the Quality of End Products*. Proc. of the 13th World Congress of Food Science and Technology (IUFOST'2006), 2006.

Description détaillée :

1. Problématique et objectifs scientifiques

Le travail proposé se situe dans la problématique générale de la prise de décision en présence d'informations contradictoires. L'objectif scientifique concerne plus spécifiquement la prise de décision en présence de points de vue contradictoires. Ainsi, dans le domaine d'application considéré, la définition de la qualité alimentaire doit concilier différentes facettes constituées par les qualités nutritionnelles, organoleptiques et sanitaires des aliments. A titre d'illustration, un exemple simple est celui de l'augmentation de la teneur en fibres d'un produit céréalier, qui peut être bénéfique d'un point de vue nutritionnel mais néfaste d'un point de vue organoleptique.

Par ailleurs, l'importance accordée à ces différents critères varie en fonction des acteurs considérés. Ainsi les experts peuvent estimer un niveau de risque lié à un contaminant comme tout à fait acceptable au vu du bénéfice apporté (par exemple pour un produit phytosanitaire) ou comparé au coût de précautions supplémentaires peu efficaces (par exemple pour une mycotoxine), tandis que le consommateur n'acceptera pas l'existence d'un risque alimentaire même minime.

Le questionnement scientifique sous-jacent est le suivant :

- Comment représenter, au sein d'un modèle de représentation des connaissances, ces points de vue contradictoires ?
- Comment représenter la notion de gravité, la notion de certitude liées à un risque (ou à un bénéfice) ?
- Comment prendre en compte les priorités des différents acteurs concernés et l'importance relative qu'ils accordent aux critères considérés ?
- Peut-on imaginer des scénarios différents représentatifs des préoccupations des différents acteurs concernés ?
- Comment résoudre les conflits posés pour parvenir à un compromis au sein d'un système automatique d'aide à la décision ?

La question considérée est donc celle de l'introduction, au sein d'un modèle de représentation des connaissances, de méthodes de raisonnement en présence d'informations contradictoires et en particulier de points de vue contradictoires.

Les résultats seront appliqués et validés sur les connaissances recueillies auprès des acteurs des filières céréales, fruits et légumes et vin, incluant chercheurs, industriels mais aussi consommateurs.

2. Situation du sujet, enjeux et perspectives

La définition de la qualité d'un produit alimentaire repose sur de nombreux critères, qui sont classiquement déclinés en trois grands types de qualités constituant le triptyque qualité nutritionnelle / qualité organoleptique / qualité sanitaire. Or ces trois grandes familles de qualités et leurs différentes composantes ne sont pas toujours compatibles et leur amélioration simultanée s'avère être un problème qui n'admet pas de solution simple.

Ainsi, la consommation de produits céréaliers complets ou l'inclusion de la peau des fruits et légumes, si bénéfiques soient-elles d'un point de vue nutritionnel par leur apport en micro-nutriments et fibres, posent la question du risque de contamination, par exemple par les phytosanitaires. Faut-il choisir le bio ou se fier aux réglementations, peut-on tout sacrifier à la sécurité et à la santé au risque de sous-estimer le plaisir gustatif, le dilemme se pose pour les consommateurs, mais concerne également les acteurs des filières et les décideurs.

Le sujet proposé s'appuie sur une approche centrée sur le domaine de la représentation des connaissances et du raisonnement sur ces connaissances afin de permettre, d'une part, la formalisation des connaissances disponibles en tant qu'éléments de prise de décision, incluant l'expertise « implicite » et pas uniquement les données analytiques plus classiquement utilisées, d'autre part, l'aide à la décision par le développement de méthodes allant jusqu'à la résolution de conflits entre connaissances contradictoires.

Le domaine d'application concerne l'évaluation de la balance bénéfices/risques au sein de trois filières agroalimentaires, les céréales, les fruits et légumes et le vin. Ce triple domaine d'application doit permettre la validation des méthodes développées et démontrer leur capacité de généralisation. En effet les filières considérées diffèrent et se complètent par leurs problématiques : voie sèche ou liquide, complexité différente en termes de nombre d'opérations de transformation, points critiques différents (fractionnement, stabilisation thermique, intervention du vivant, ...).

Un compromis entre qualité nutritionnelle, organoleptique et sanitaire s'est construit de manière empirique au sein des filières, avec la maîtrise progressive des procédés de transformation. L'émergence de nouvelles préoccupations et de nouvelles demandes implique aujourd'hui le déplacement de ce compromis vers un nouvel équilibre, faisant en particulier une plus large place aux aspects nutritionnels.

Ainsi les politiques de santé publique, dont fait partie le PNNS lancé en 2001, tentent de faire face à certaines maladies en augmentation dans nos sociétés occidentales (maladies cardiovasculaires, cancers, obésité, ...). Le coût annuel de l'obésité et des maladies associées a été évalué à 23 milliards d'euros en Italie, tandis qu'en France, l'obésité chez les enfants augmente de façon spectaculaire (13.3 % en 2000).

Les consommateurs se sensibilisent à ces nouvelles problématiques et les industriels ont besoin d'outils pour répondre aux demandes émergentes par l'adaptation, l'innovation, l'optimisation des procédés de transformation au sein des filières.

3. Description de la méthodologie

Choix du formalisme de représentation des connaissances

Un élément essentiel dans le choix du langage de représentation des connaissances utilisé est qu'il soit suffisamment intuitif pour les utilisateurs : les connaissances représentées, mais aussi les raisonnements effectués, doivent être aisément compréhensibles et visualisables. Le formalisme choisi doit également permettre la représentation de connaissances symboliques ainsi que de concepts ayant une organisation hiérarchique. Enfin il doit être suffisamment souple pour pouvoir être étendu aux objectifs du travail.

La plupart des formalismes de représentation de connaissances actuels sont basés sur la logique, en particulier la logique du premier ordre pour les formalismes les plus connus – les logiques de description – difficilement accessible pour un utilisateur non spécialiste.

Nous proposons donc d'utiliser un formalisme basé sur des diagrammes, plus précisément sur les graphes conceptuels, introduits par Sowa en 1984. Les propriétés graphiques du modèle, tant au niveau de la représentation qu'au niveau des mécanismes de raisonnement, facilitent l'interprétation intuitive par l'utilisateur des connaissances mais aussi des raisonnements effectués. Toutefois à la différence de nombreux modèles graphiques utilisés en ingénierie des connaissances, les graphes conceptuels possèdent de plus une interprétation logique, et les raisonnements graphiques ont une équivalence en déduction logique, ce qui permet de fonder formellement ces raisonnements. Enfin, la mise en œuvre effective des raisonnements sur une base de connaissances de grande taille est un aspect important. Les mécanismes étant basés sur des opérations de théorie des graphes, leur implémentation bénéficie des nombreux travaux algorithmiques liés à ce domaine (comme ceux sur les réseaux de contraintes notamment).

Existant et extensions prévisibles du formalisme

Le modèle des graphes conceptuels simples permet de représenter des connaissances factuelles en s'appuyant sur un vocabulaire qui inclut un ensemble de concepts organisés de façon hiérarchique. L'opération fondamentale, appelée projection, est un appariement de graphes qui peut être utilisé comme mécanisme de recherche d'information.

Plusieurs extensions du modèle de base offrent un intérêt particulier pour le projet :

- des connaissances plus complexes peuvent être exprimées sous la forme de règles d'inférence « si ... alors... » et de contraintes sur le domaine d'application considéré ;
- les graphes conceptuels emboîtés offrent une manière de représenter la notion de point de vue, en permettant de préciser la portée de l'assertion d'une connaissance particulière ;
- les graphes conceptuels flous permettent d'exprimer des préférences sur un ensemble de concepts, ou des connaissances imprécises dans lesquelles plusieurs concepts candidats sont envisagés.

Ces éléments existants nous font penser qu'il est possible d'adapter et de développer le modèle des graphes conceptuels pour atteindre les objectifs visés par le sujet proposé.

Toutefois un important travail de recherche est nécessaire, dans la mesure où :

- les différentes extensions citées ci-dessus n'ont jamais été combinées en un modèle unique qui offrirait alors une grande richesse de représentation ;
- la représentation de points de vue actuellement permise par les graphes conceptuels emboîtés reste extrêmement restreinte par rapport aux besoins du projet. En particulier, il est difficile de faire cohabiter plusieurs natures de points de vue (par exemple par type de qualité considérée nutritionnelle/organoleptique/sanitaire et par type d'acteur chercheur/industriel/consommateur). De plus les conflits entre points de vue ne sont pas gérés ;
- la représentation de préférences et de connaissances imprécises actuellement existante ne permet pas le niveau d'expressivité nécessité par le projet. Ainsi, pour des raisons de clarté

sémantique, il est impossible d'exprimer dans un même graphe à la fois des préférences d'acteurs, des niveaux de gravité de risque et des niveaux d'incertitude : le formalisme utilisé étant homogène (sous-ensembles flous), il y aurait confusion entre les trois types d'interprétation ;

- l'apport d'autres domaines, en particulier les travaux sur la fusion de connaissances, et l'expression de préférences, devrait être intégré à différents niveaux du modèle (notamment au niveau des emboîtements) tandis qu'il n'est actuellement pris en compte qu'au niveau de la hiérarchie des concepts dans les graphes conceptuels flous ;
- le déclenchement de règles différentes selon le niveau de granularité considéré reste à définir ;
- enfin une difficulté essentielle de ce travail consiste à proposer une extension du modèle qui conserve ses propriétés graphiques et une équivalence en logique.

Au niveau informatique, il s'agit donc, en s'appuyant sur nos résultats théoriques, algorithmiques, logiciels et applicatifs dans le domaine des graphes conceptuels, d'adapter, d'étendre et d'intégrer d'autres résultats obtenus dans diverses problématiques de l'intelligence artificielle, tout en conservant un langage graphique, fondé d'un point de vue logique, et muni de mécanismes de raisonnements mis en œuvre par des algorithmes suffisamment efficaces pour pouvoir être utilisés sur des applications réelles.

Lien avec le domaine d'application

Concernant la partie applicative, le recueil de connaissances – incluant bibliographie, dires d'experts, données expérimentales disponibles (il n'est pas prévu d'expérimentations), enquêtes – portera notamment (mais pas exclusivement) sur les éléments suivants :

- les aspects nutritionnels et contaminants des enveloppes (blé) et de la peau (tomate), éléments de comparaison entre les deux filières, mais aussi abordant la question de constituants positifs et négatifs co-localisés ;
- dans la filière vin, l'impact de techniques de vinification impliquant un chauffage (thermovinification, macération pelliculaire à chaud, flash détente), qui d'une part favorisent l'extraction et la stabilisation de la couleur ainsi que l'apparition d'arômes fruités, mais d'autre part dégradent certains résidus phytosanitaires en métabolites plus toxiques (mancozène en ETU) ou diminuent les teneurs en certains nutriments comme les vitamines B (qui pourraient jouer un rôle dans la lutte contre l'addiction à la consommation d'alcool).

Les connaissances recueillies (évolution au long de la filière, impact des procédés, réglementations et recommandations, moyens d'action, perception bénéfiques/risques par les différents acteurs, ...) seront formalisées dans un outil informatique, « traduites » dans le modèle des graphes conceptuels.

4. Calendrier

Année 1 :

- bibliographie
- état de l'art dans le domaine de l'aide à la décision
- évaluation des extensions existantes du modèle des graphes conceptuels

Année 2 :

- propositions sur la cohabitation de plusieurs natures de points de vue (acteurs, objectifs)
- propositions sur la représentation de plusieurs échelles de priorités (préférences, gravité/bénéfice en jeu, certitude du risque/bénéfice)

Année 3 :

- propositions sur la résolution de conflits
- synthèse et rédaction

5. Collaborations envisagées

Outre la collaboration entre l'UMR IATE (pluri-disciplinaire technologies des céréales / représentation des connaissances) et le LIRMM, des collaborations sont prévues d'une part au niveau théorique avec l'équipe RPDMP (Raisonnements Plausibles, Décision, Méthodes de Preuves) de l'IRIT (Institut de Recherche en Informatique de Toulouse), d'autre part au niveau applicatif avec des partenaires des filières fruits et légumes (UMR SQPOV Avignon) et vin (UMR SPO, ITV).