

Équipe DECIDE du Lab-STICC

Contact

Patrick Meyer patrick.meyer@telecom-bretagne.eu
 John Puentes john.puentes@telecom-bretagne.eu
 Télécom Bretagne, Brest

Introduction

L’ambition de l’équipe DECIDE est de fournir des solutions qui permettent de faciliter la prise de décision lorsque un ou plusieurs décideurs sont confrontés à des données hétérogènes et complexes. Les recherches menées au sein de l’équipe DECIDE² couvrent toutes les étapes clés permettant de passer des données à la décision. Elles débutent en amont du processus de décision avec la formalisation du problème et la récolte des données, et finissent en aval avec la validation des recommandations de décision. Elles passent par l’extraction, la qualification, l’apprentissage et la modélisation des données, des valeurs, des préférences, l’étude des aspects algorithmiques sous-jacents jusqu’à la conception des méthodes et systèmes de soutien à la prise de décision. La recherche menée au sein de DECIDE est appliquée à des thématiques extrêmement variées comme la planification de levés hydrographiques, l’assistance aux personnes dépendantes, l’évaluation économique de technologies de santé, le contrôle de drones, l’indexation d’images, l’aide au codage médical, l’analyse automatique de textes, ...

DECIDE est une équipe de l’UMR 6285 Lab-STICC, et se compose d’une trentaine de personnes, parmi lesquelles 9 enseignants-chercheurs, 1 ingénieur de recherche CNRS et 1 ingénieur de recherche Télécom Bretagne. Les membres de l’équipe sont issus de plusieurs institutions : Télécom Bretagne, l’Université de Bretagne Occidentale, l’Université de Bretagne Sud et l’ENSTA Bretagne. 10 thèses de doctorat ont été soutenues au cours de la période 2010-2015, ainsi qu’une HDR. A la date du 30 juin 2015, 8 doctorants effectuent leurs travaux au sein de l’équipe.

2. <http://www.labsticc.fr/le-pole-cid/decision-aid-and-knowledge-discovery-decide/>

Axes de recherche de l’équipe

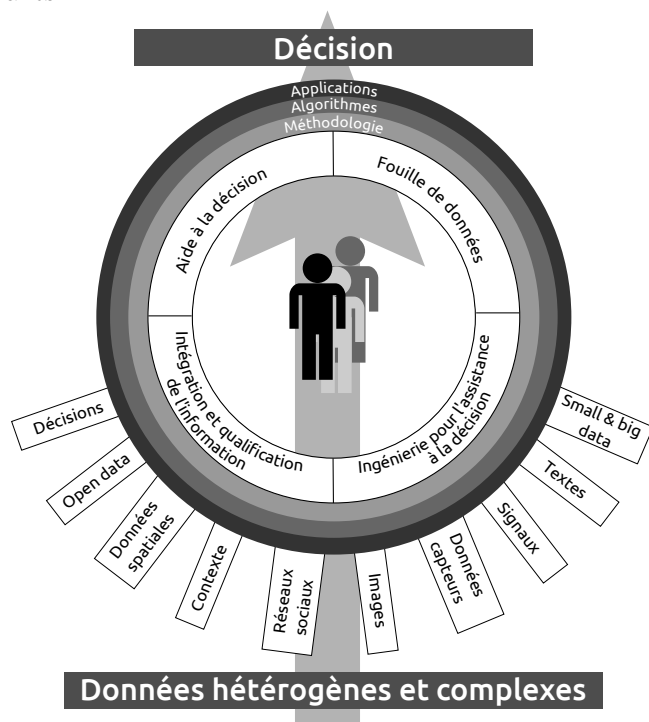
Les données hétérogènes constituent le point de départ de la recherche menée au sein de l’équipe DECIDE. Elles peuvent être constituées de textes, de signaux, d’images, de données issues de capteurs, d’informations contenues dans des réseaux sociaux, de la description du contexte de la prise de décision, de préférences de décideurs, de données spatiales, ou même des décisions prises précédemment. Le ou les acteurs qui sont confrontés à ces données complexes peuvent se retrouver face à la difficulté de prendre une décision éclairée. Il est donc important de créer un lien entre ces données et la prise de décision, en proposant des solutions qui permettent au(x) décideur(s) :

- d’identifier, de modéliser et de comprendre l’information contenue dans les données ;
- de prendre des décisions fiables et robustes à partir de cette information ;
- de justifier la recommandation de décision afin d’en attester la qualité ;
- d’accompagner le décideur vers une recommandation acceptable à ses yeux ;
- de conserver la trace du processus qui a mené à la décision en vue de garantir une certaine lisibilité et sa traçabilité ;
- d’identifier leurs préférences, potentiellement incertaines ;
- d’atteindre un compromis entre plusieurs décideurs aux préférences conflictuelles.

La recherche menée au sein de DECIDE ambitionne en conséquence à répondre à ces différents défis liés à la prise de décision.

Afin de consolider ce lien entre les données hétérogènes et la décision, les travaux de recherche de l’équipe se déclinent selon 4 axes de recherche, notamment, l’aide à la décision, la fouille de données, l’intégration et la qualification de l’information et l’ingénierie pour l’assistance à la décision. La figure ci-après présente du bas vers le haut ce passage des données à la décision. Les techniques génériques utilisées de manière commune par les quatre axes sont principalement la théorie des graphes, les techniques d’optimisation, l’apprentissage automatique, la visualisation et la modélisation. L’ensemble des axes contribue à concevoir des nouveaux algorithmes et outils adaptés à

des applications spécifiques, avec comme objectif de fournir une assistance à la prise de décision. Dans ce schéma on retrouve également les acteurs humains du processus d'assistance, qui sont centraux dans les outils et résultats produits par l'équipe DECIDE, tout comme les décideurs. Ces axes ainsi que leurs sous-thèmes permettent à l'équipe de couvrir quatre aspects importants de l'informatique décisionnelle. Dans la suite nous détaillons ces quatre axes.



Axe “aide à la décision”

Dans l'équipe DECIDE, l'aide à la décision vise à soutenir un ou plusieurs individus, confrontés à des alternatives de décision décrites par des conséquences multiples, à prendre une décision en leur fournissant des recommandations. Cet axe est actuellement divisé en 4 thèmes : la modélisation et l'élicitation des préférences, l'aide multicritère à la décision, les systèmes de recommandation, et les systèmes d'information géographiques.

Axe “fouille de données”

La fouille de données concerne l'extraction des connaissances à partir de grandes quantités de données (structurées et partiellement structurées) par des méthodes automatiques ou semi-automatiques. On retrouve dans cet axe les 5 thèmes suivants : la recherche de motifs fréquents et réguliers, la classification supervisée, la fouille

de texte et le traitement automatique de la langue, la fouille de graphes, et, pour terminer, la visualisation, avec notamment les difficultés liées à la synthèse de grands volumes de données.

Axe “intégration et qualification de l'information”

L'axe d'intégration et qualification de l'information se focalise sur l'agrégation de données, ayant comme objectif de créer des indicateurs en intégrant des données provenant de sources hétérogènes, afin de qualifier leur pertinence dans un contexte particulier. L'intérêt de l'équipe porte considérablement sur des mesures de qualité de l'information fusionnée, la pertinence des descripteurs du contenu visuel, la qualité de l'information en accord avec l'usage, ainsi que la caractérisation et la généralisation de propriétés algorithmiques de mesures de qualité.

Axe “ingénierie pour l'assistance à la décision”

Cet axe contient essentiellement des outils au sens large, qui permettent de faciliter le processus d'assistance au décideur. Il est composé de 2 thèmes : les plateformes informatiques et l'ingénierie dirigée par les modèles pour la décision. Les plateformes représentent les implémentations informatiques des outils développés par l'équipe, alors que l'ingénierie dirigée par les modèles pour la décision traite la complexité qualitative ou quantitative des données en entrée du processus de décision. La démarche est d'abord de modéliser un problème de décision par les concepts qui y participent et par leurs relations. Ensuite, la solution conçue pour le problème est implémentée par une, ou plusieurs, transformations du modèle de données.

D'autre part, il existe des liens émergents entre les quatre axes. Ainsi, les résultats de la fouille de données, de l'intégration et de la qualification des données peuvent servir comme point de départ pour l'aide à la décision, et en particulier la modélisation des préférences. De même, les modèles des préférences des décideurs ou l'ingénierie dirigée par les modèles peuvent servir pour améliorer l'efficacité des algorithmes de fouille, d'intégration et de qualification de données, ou tout simplement être utiles dans une phase de prétraitement des données.

Dans son projet, l'équipe cherche à développer et à amplifier les liens entre les 4 axes et à stimuler les synergies entre eux. À cet égard, diverses interdépendances émergentes entre ces thématiques indiquent clairement que chacune est potentiellement en mesure d'appliquer, d'adapter et de faire évoluer les méthodes des autres thématiques, afin de renforcer l'originalité de la production

de l'équipe et d'élargir ses axes de compétence. Les plateformes informatiques viennent évidemment soutenir les différents axes lors des applications.

Chacun de ces axes est par ailleurs traité via trois perspectives. Tout d'abord le point de vue “méthodologie” autrement dit la définition d'approches méthodologiques adaptées aux problèmes traités et basées sur les résultats produits par les 4 axes. Dans la perspective “algorithmique”, il s'agit de développer des algorithmes de résolution originaux et de faire évoluer des solutions existantes, notamment en vue d'alimenter la recherche méthodologique. Pour terminer, les “applications”, ont traité à tester la validité d'algorithmes, d'outils et de méthodologies, et de les appliquer en vue de résoudre des problèmes de décision réels.

Ces trois points de vue garantissent une vision globale du passage des données à la prise de décision. En effet, les problèmes réels viennent alimenter la recherche fondamentale et méthodologique. Par ailleurs, les cas applicatifs contribuent également à améliorer, tester et valider l'outillage. Finalement, la confrontation des travaux théoriques à des problèmes de décision réels permet leur validation de terrain.

Bénéficiant d'une structure pluridisciplinaire, l'équipe se positionne plus particulièrement sur les thématiques suivantes :

- Modélisation et opérationnalisation du processus d'aide multicritère à la décision et outillage.
- Caractérisation de la qualité de l'information de plusieurs systèmes interconnectés, basée sur l'utilisation du système par des experts.
- Caractérisation et généralisation des propriétés de mesures d'intérêt de règles.
- Intégration d'attributs dans l'analyse de réseaux complexes.
- Modélisation fonctionnelle des services du SI conformément à un modèle métier.
- Traitement automatique de la langue en combinaison avec les langages visuels (de type carte géographique ou figures géométriques) pour former des langages hybrides.

Quelques faits marquants de ces dernières années

- Création du Master of Science (MSc) in Computer Science & Decision Systems (CSDS) (2010).
- Participation aux chaires “Cyberdéfense des Systèmes Navals” (Télécom Bretagne, Ecole Navale, DCNS, Thales) (1 thèse, 1 post-doc) et “Réseaux sociaux” (Ins-

titut Mines-Télécom, La Poste, Pages Jaunes Group, Danone) (1 post-doc).

- Création et coordination de la Task Force “Evaluation and Quality” (IEEE Data Mining and Big Data Analytics Technical Committee) (2011).
- Co-édition d'un ouvrage de référence en Aide Multicritère à la Décision : *R. Bisdorff, L. Dias, P. Meyer, V. Mousseau, M. Pirlot, Evaluation and Decision Models with Multiple Criteria : Case Studies. Springer, juin 2015.*
- La plateforme diviz d'Aide Multicritère à la Décision est devenue un outil de référence dans la communauté (plus de 750 utilisateurs à travers le monde, plusieurs milliers de traces d'utilisations par an).

Publications

Pour la période 2010-2015, les publications de l'équipe englobent, 52 articles de revue avec comité de lecture, 4 articles de revue sans comité de lecture, 12 chapitres dans un livre, 106 communications dans une conférence avec acte, 37 communications dans une conférence sans acte, 2 conférences invitées, 5 directions d'ouvrage, 1 HDR et 5 rapports techniques.

Sélection de publications représentatives récentes

- Simonin, J., Nurcan, S., and Barrios, J. (2014). Méthode automatisée de conception de scénarios d'évolution organisationnelle fondée sur les processus et les buts. *Ingénierie des systèmes d'information*, 19(2) : 9 – 33.

Cet article souligne l'intérêt théorique d'une vue stratégique de l'entreprise, nouvelle vue d'architecture d'entreprise, décrivant ses buts et leurs relations. L'alignement de la vue métier de l'entreprise (organisation et processus métiers) sur la vue stratégique permet de proposer automatiquement des scénarios d'évolution de l'organisation facilitant l'atteinte des buts. Cette application de l'ingénierie des modèles est aussi considérée comme un complément significatif par la communauté.

- Puentes, J., Montagner, J., Lecornu, L., and Cauvin, J. M. (2013). Information quality measurement of medical encoding support based on usability. *Computer methods and programs in biomedicine*, 112(3) : 329 – 342.

Cet article définit un nouveau modèle de mesure de la qualité d'information d'un ensemble de taille variable de codes médicaux, assemblés automatiquement par un système d'aide au codage. L'article montre que

les méthodes utilisées dans la littérature pendant les 30 dernières années ne sont pas adaptées et pourquoi elles sont limitées. La méthode proposée se base sur la manière de travailler des codeurs et peut facilement comparer des centaines de codes.

- Cruz Gomez, J. D., Bothorel, C., and Poulet, F. (2013). Community detection and visualization in social networks : integrating structural and semantic information. *ACM transactions on intelligent systems and technology (TIST)*, 5(1) : art.no11.

Cette publication démontre que le domaine de l'Intégration d'attributs dans l'analyse de réseaux complexes est balbutiant et positionne l'équipe DECIDE parmi les pionniers, en proposant un modèle et un algorithme pour résoudre le problème d'intégration d'attributs dans le clustering de graphes avec attributs mais aussi un layout de visualisation permettant une analyse de bout en bout, des données brutes hétérogènes jusqu'à l'analyse visuelle pour déterminer les acteurs clés et comprendre les interactions entre communautés sociales thématiques.

- Le Bras, Y., Lenca, P., and Lallich, S. (2012). Optimotone measures for optimal rule discovery. *Computational Intelligence*, 28(4) : 475 – 504.

Cet article généralise la propriété d'optimalité des règles et en propose une caractérisation mathématique (une règle n'est pas optimale si elle a un intérêt plus faible qu'une de ses généralisations). La propriété d'optimotone ainsi proposée permet de développer des algorithmes dirigés par une heuristique permettant d'élaguer l'espace de recherche de façon efficace et de retourner l'ensemble complet des règles optimales pour une mesure optimotone.

- Cailloux, O., Meyer, P., and Mousseau, V. (2012). Eliciting ELECTRE TRI category limits for a group

of decision makers. *European journal of operational research*, 233(1) : 133 – 140.

Cet article propose des outils algorithmiques et un processus qui permet d'aboutir à un modèle des préférences unique et consensuel pour un groupe de décideurs. Il répond au problème délicat de l'élicitation des préférences d'un groupe de décideur et propose des outils opérationnels pour mettre en œuvre la méthodologie proposée.

- Yannis Haralambous, Philippe Lenca, Text Classification Using Association Rules, Dependency Pruning and Hyperonymization, *Data Mining and Natural Language Processing* (workshop de ECML-PKDD), Nancy 2014 (CEUR-WS, vol. 2012).

Cet article combine des approches jusqu'ici spécifiques au traitement automatique de la langue, telles que les dépendances syntaxiques (selon la théorie syntaxique de Tesnière) et l'hypéronymisation (c'est-à-dire le remplacement d'un mot par ses hypéronymes extraits de ressources de type WordNet), à des algorithmes de classification tels que les SVM et les règles d'association de classes. En agrégeant différentes méthodes de sélection de descripteurs et différents types de classifieurs nous avons pu optimiser la tâche de classification de textes de langue anglaise.

Brest, le 30 juin 2015

Cécile Bothorel
Yannis Haralambous
Philippe Lenca
Patrick Meyer
John Puentes
Jacques Simonin

Équipe GraphIK (Graphs for Inferences on Knowledge)

Contact

Croitoru Madalina croitoru@lirmm.fr

Buche Patrice patrice.buche@supagro.inra.fr

Établissements : INRIA (Sophia Antipolis Méditerranée center), the University of Montpellier 2, CNRS and INRA.

<https://team.inria.fr/graphik/>

GraphIK (Graphs for Inferences on Knowledge) is a joint team between INRIA (Sophia Antipolis Méditerranée center), the University of Montpellier 2, CNRS and INRA. It has been created in 2010.

In GraphIK we currently focus on two application domains : knowledge representation in agronomy (more precisely applied to the quality in agri-food chains), and metadata management (in particular for bibliographic metadata). The application to agronomy is determined