

Vers un outil électronique pour aider à l'identification et à la reconnaissance des essences de bois

BAUDRIT Cédric¹, MICHAUD Franck², FERNANDEZ Christophe³

^{1,3} INRAE, Université de Bordeaux-I2M, USC 1368, F-33400 Talence, France

² LIMBHA, Ecole Supérieure du Bois, 7 Rue Christian Pauc, F-44306 Nantes, France

cedric.baudrit@inrae.fr

Mots clefs : Modèle graphique probabiliste ; machine learning ; anatomie du bois ; approche macroscopique

Contexte et objectifs

La reconnaissance des essences de bois dans un morceau de bois ou un objet en bois et l'identification des caractéristiques anatomiques de ces essences sont des problèmes difficiles dans la mesure où le bois est un matériau complexe et hétérogène (Corbineau et Macchioni 2015). La reconnaissance et/ou l'identification d'une essence de bois, à l'échelle d'une pièce, nécessite de pouvoir manipuler, agréger et intégrer un certain nombre de variables hétérogènes simultanément. S'il existe déjà des outils pour la reconnaissance des essences, ils sont toutefois principalement basés soit sur de l'analyse d'image ou mesures physiques (approche boîte noire de l'identification (Wielenzynski 2016), soit sur le principe d'élimination par filtre (approche de sélection sans véritable prise en compte de l'utilisateur et des spécificités ou difficultés de la démarche en pratique), soit ce sont des outils de spécialiste qui nécessitent un niveau de connaissances très élevé (chercheur du domaine). Il est donc nécessaire de proposer un outil de simulation capable de mutualiser des connaissances hétérogènes dans le but (1) de guider, en situation de pratique professionnelle sur site, un utilisateur, un apprenant ou un expert dans l'identification d'une essence au travers d'un ensemble de questions et/ou (2) de fournir des conclusions et des propositions pertinentes de descripteurs, caractéristiques les plus vraisemblables d'une essence (re)connue (Fig. 1). Dans ce but, nous nous concentrons sur l'utilisation du concept de réseaux bayésiens (BNs) (Jensen et Nielsen 2010, Pearl 1988) qui fournit un formalisme mathématique pratique permettant de décrire des systèmes complexes entachés d'incertitude. Ils s'appuient sur les modèles graphiques probabilistes où la structure du réseau fournit une interface intuitive à travers laquelle un individu peut modéliser des ensembles de variables en interactions et fournit une représentation qualitative des connaissances. L'incertitude relative au système est prise en compte en quantifiant la dépendance entre les variables à l'aide de probabilités conditionnelles. Les BNs permettent de combiner différentes sources de connaissances expertes avec des données expérimentales à différents niveaux et échelles de connaissances. Cette approche a été étudiée et utilisée dans des domaines qui vont du biomédical à la pétro-physique (Wiegerinck et al 2013). L'objectif est donc de concevoir un outil électronique utilisable sur le terrain (ex. smartphone/tablette) pour aider à l'apprentissage et à la reconnaissance des essences de bois.

Matériel et méthode

Un réseau Bayésien (BN) (Jensen et Nielsen, 2010 ; Pearl, 1988) est un modèle graphique probabiliste représentant un ensemble de variables aléatoires en interaction sous la forme d'un graphe. Formellement, un réseau Bayésien est un Graphe orienté acyclique (DAG) dont les nœuds représentent des variables aléatoires et dont les arcs manquants codent les indépendances conditionnelles entre les variables.

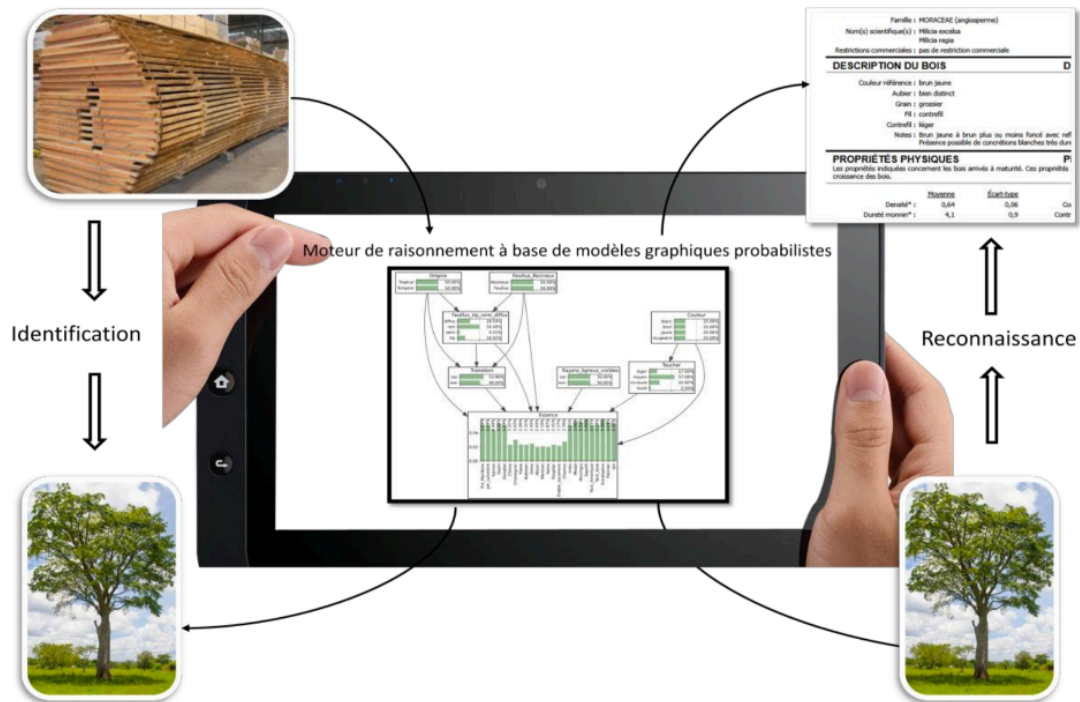


Fig. 1: outil électronique pour aider à identifier et reconnaître des essences de bois

Le graphe est appelé la structure du réseau et les nœuds contenant l'information probabiliste sont appelés les paramètres du réseau. Dans un réseau Bayésien, la distribution de probabilités jointes des valeurs des nœuds peut être écrite comme le produit de la distribution de probabilité de chaque nœud et de ses parents. Si l'ensemble des nœuds parents d'un nœud X_i est désigné par $Pa(X_i)$, nous avons :

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa(X_i)) \quad (\text{Eq.1})$$

où $P(X_1, \dots, X_n)$ représente la probabilité jointe des variables X_1, \dots, X_n et $P(X_i | Pa(X_i))$ représente la probabilité conditionnelle associée à la variable X_i sachant $Pa(X_i)$. Différentes techniques d'apprentissage permettent d'identifier soit la structure du graphe (*i.e* la topologie), soit les paramètres du réseau (*i.e* les distributions de probabilités conditionnelles) ou une combinaison des deux à partir de données substantielles et/ou incomplètes combinées à une élicitation par des experts (Buntine 1996, Heckerman 2008, O'Hagan 2006). Ji et al. (2015) donnent un aperçu des méthodes d'apprentissage des paramètres dans les réseaux bayésiens. L'utilisation de ces réseaux consiste en une "requête" exprimée sous forme de probabilités conditionnelles. La tâche la plus courante consiste à estimer les probabilités marginales $P(X_Q | X_E)$ où X_Q est un ensemble de variables de requête et X_E un ensemble de variables observées. L'inférence consiste alors à calculer la probabilité de chaque état d'une variable X_Q lorsque nous connaissons l'état pris par les autres variables X_E . Pour plus de détails sur l'inférence, le lecteur peut se référer à Guo et Hsu (2002), Hrycej (1990), Salmeron *et al.* (2000), Madsen et Jensen (1999) qui présentent différents types d'algorithmes d'inférence (inférence exacte et approximative) selon la complexité et la taille du réseau (Cooper 1990).

Résultats et discussion

La structure du réseau a été établie par expertise pour mener 2 types de raisonnement à savoir (1) proposer l'essence de bois la plus vraisemblable qui compose l'objet en bois observé en répondant à un questionnaire et (2) proposer l'ensemble des descripteurs les plus

vraisemblables face à une essence reconnue. Chaque nœud représente une caractéristique des essences et les paramètres du réseau ont été estimés à partir de la connaissance experte et d'une base de données composée de 25 essences de bois. Formellement, nous allons estimer les probabilités suivantes :

$$\text{Proba}(\text{Essence} \mid \text{Descripteurs}) \quad \text{et} \quad \text{Proba}(\text{Descripteurs} \mid \text{Essence reconnue})$$

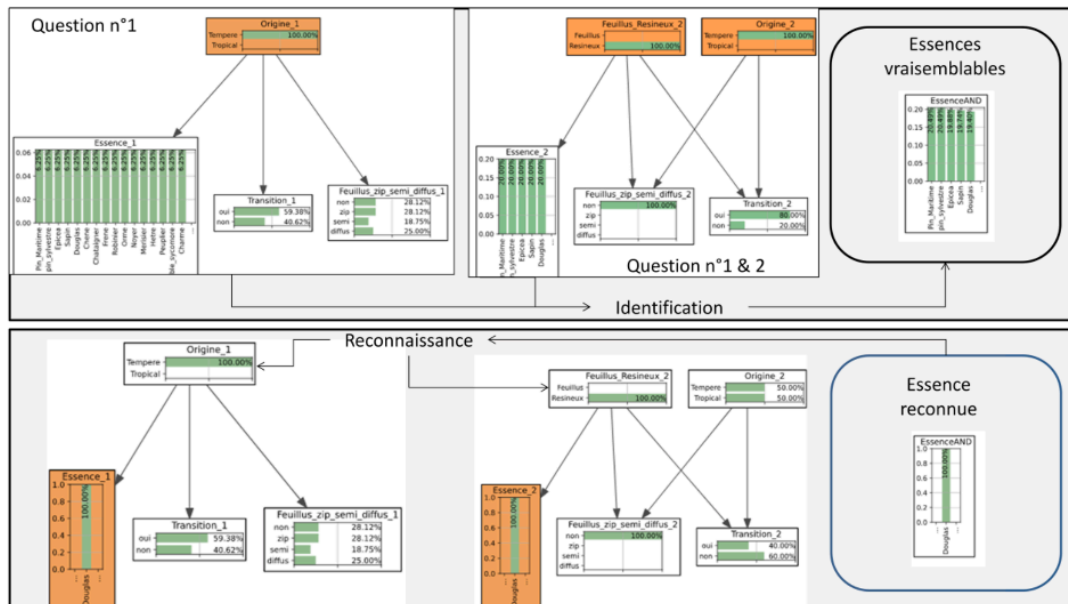


Fig. 2 : Simulation du raisonnement pour (1) identifier une essence de bois à partir d'un questionnaire (haut) et (2) reconnaître les caractéristiques du Douglas ou *Pseudotsuga menziesii* (bas)

Le modèle a été construit et implémenté en utilisant la bibliothèque python pyAgrum basée sur la bibliothèque C++ aGrUM (Gonzales et al., 2017) dédiée aux modèles graphiques probabilistes permettant de créer, gérer et effectuer des inférences avec les réseaux bayésiens. La figure 2 haute présente par exemple la construction du modèle à partir des réponses aux 2 questions suivantes : Quelle est l'origine géographique de l'essence de bois utilisée (question n°1) ? S'agit-il d'un résineux ou d'un feuillus (question n°2) ? L'utilisateur répond par exemple à la première question par : « le bois utilisé provient d'une zone tempérée » (ie $\text{Proba}(\text{Origine}_1=\text{Tempere})=1$ dans le modèle, voir Fig. 2) ; le modèle élimine ainsi toutes les essences de bois tropicales et donne une indication sur la transition abrupte ou progressive entre les deux parties du cerne (transition entre le bois initial et le bois final). Le modèle estime ainsi qu'à 59.38% la transition sera vraisemblablement progressive (ie $\text{Proba}(\text{Transition}_1=\text{oui}|\text{Origine}_1=\text{Tempere})=59.39$). Cette vraisemblance est renforcée à 80% en répondant à la question 2 par le fait que l'essence de bois fait partie des résineux (ie $\text{Proba}(\text{Transition}_1=\text{oui}|\text{Origine}_1=\text{Tempere}, \text{Feuillus_Resineux}_2=\text{Résineux})=80\%$). Le fait d'observer finalement une transition brutale par la suite pourrait remettre en cause l'origine du bois par exemple. Le modèle ne propose finalement à partir des réponses aux questions 1 & 2 que (voir la variable EssenceAND Fig. 2) 5 choix possibles d'essences sans pouvoir les discriminer. Ceci indique qu'il faut aller plus loin dans le questionnaire pour réduire le nombre de choix. La partie inférieure de la Fig.2 présente la cartographie des caractéristiques les plus vraisemblables face à une essence reconnue à savoir le Douglas. Le modèle estime que le Douglas est un résineux d'origine tempérée, (ie $\text{Proba}(\text{Origine}_1=\text{Tempere}|\text{essence}=\text{Douglas})=1$ & $\text{Proba}(\text{Feuillus_Resineux}_2=\text{Resineux}|\text{essence}=\text{Douglas})=1$).

Conclusion et perspectives

Un modèle générique à base de graphes probabilistes a été implémenté permettant à la fois de proposer (1) l'essence la plus vraisemblable qui a été utilisée ou présente dans un morceau de bois ou un objet en bois et (2) un ensemble de descripteurs vraisemblables qui caractérise une essence reconnue. Les descripteurs choisis sont des observables à l'échelle macroscopique (ou un grossissement maximal de x10), plus ou moins facilement accessibles (selon la qualité de la pièce et de l'état de surface), qui s'inscrivent dans la démarche pratique et pragmatique d'apprentissage de l'identification et la reconnaissance des essences de bois déployée depuis 2014 à l'École Supérieure du Bois.

Remerciements

Ces travaux sont en partie financés par le programme de coopération franco-québécois CHAMPLAIN

Références

- Buntine, W. (1996) A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on Knowledge and Data Engineering* 8, 195–210.
- Cooper, G. F. (1990) The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial intelligence*, 42(2-3), 393-405.
- Corbineau P., Macchioni N. (2015) *Connaitre, Reconnaître et Nommer le Bois « Savoir&Faire : le bois »* sous la direction d'Hugues Jacquet.
- Gonzales C., Torti L., Wuillemin P. (2017) *Agrum: a graphical universal model framework*. In *Int. Conference on Industrial, Engineering, Other Applications of Applied Intelligent Systems*, Arras, France. Springer-Verlag.
- Guo H., Hsu W. (2002) A survey of algorithms for real-time Bayesian network inference. In *Join Workshop on Real Time Decision Support and Diagnosis Systems*.
- Heckerman D. (2008) A Tutorial on Learning with Bayesian Networks, In *Innovations in Bayesian networks* (pp. 33–82), Springer, Berlin, Heidelberg.
- Hrycej T. (1990) Gibbs sampling in Bayesian networks, *Artif. Intell. Rev.*, 46(3), 351-363.
- Jensen F.V., Nielsen, T.D. (2010) *Bayesian Networks and Decision Graphs*. Springer.
- Ji Z., Xia, Q., Meng G. (2015) A review of parameter learning methods in Bayesian network. In *International Conference on Intelligent Computing* (pp. 3-12). Springer, Cham.
- Madsen A. Jensen F. (1999) Lazy propagation: a junction tree inference algorithm based on lazy evaluation, *Artificial Intelligence*, 113(1-2), 203–45.
- O'Hagan A. (2006) *Uncertain Judgements: Eliciting Experts' Probabilities*, Wiley, New York.
- Pearl J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Diego, 552p.
- Salmeron A., Cano A., Moral S. (2000) Importance sampling in Bayesian networks using probability trees, *Computational Statistics and Data Analysis*, 34(4), 387–413.
- Wiegerinck W., Burgers W., Kappen B. (2013) Bayesian networks, introduction and practical applications. In *Handbook on Neural Information Processing* (pp. 401-431). Springer, Berlin, Heidelberg.
- Wielezynski F. (2016) Rapport du REdEME, Rapport CODIFAB REF. LBO/DR/403/16/207 – 07/06/2016.