HAL
open science

Intégration d'un réseau bayésien dans une ontologie

Emna Hlel, Salma Jamoussi, Abdelmajid Ben Hamadou

To cite this version:

Emna Hlel, Salma Jamoussi, Abdelmajid Ben Hamadou. Intégration d'un réseau bayésien dans une ontologie. Catherine Faron-Zucker. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. pp.295-297. hal-01016410

HAL Id: hal-01016410

https://inria.hal.science/hal-01016410

Submitted on 30 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intégration d'un réseau bayésien dans une ontologie

Emna Hlel, Salma Jamoussi et Abdelmajid Ben Hamadou

Laboratoire MIRACL, Pôle Technologique de Sfax BP 242 - 3021, Sakiet Ezzit Sfax, TUNISIE,
emnahlel@gmail.com, jamoussi@gmail.com et abdelmajid.benhamadou@gmail.com

Résumé : L'augmentation et la diversification d'informations ont créé de nouveaux besoins utilisateurs. Les problèmes de représentation, traitement, analyse et raisonnement sur l'informations, et surtout les informations incertaines, constituent encore un thème de recherche important. Donc, il est indispensable de proposer des nouvelles approches permettant de représenter formellement les informations incertaines pour aider les machines à les comprendre et à inférer des nouvelles connaissances. Cet article est inscrit dans ce cadre.

Mots-clés : ontologie probabiliste (OP), ontologie classique, réseau bayésien (RB), inférence, incertain, etc.

1 Introduction

Selon (Studer R. et al., 1998), une ontologie est «*une spécification explicite et formelle d'une conceptualisation partagée*». Elle est une représentation qui regroupe un ensemble de concepts et relations décrivant un domaine particulier. Un de principaux défauts de l'ontologie est leur incapacité de représenter et raisonner sur l'incertitude. Dans les années précédentes, divers chercheurs (Nottelmann H. & Fuhr N., 2006), (da Costa et al.,2005), (Heinsohn J., 1994), (Pool M. & Aikin J., 2004), etc ont tenté de proposer des approches visant à intégrer l'incertitude dans les ontologies, qui est un axe de recherche intéressant sur lequel nous avons situé ce travail. L'une de caractéristiques de RB est leur capacité de représenter les informations incertaines sous la forme d'un modèle probabiliste. Mais, malheureusement cette représentation n'est pas formelle et les machines ne sont pas capables de la comprendre. Pour cette raison, nous tentons de proposer une méthode permettant de représenter formellement un RB sous la forme d'une ontologie probabiliste à l'aide d'un standard de représentation d'ontologie OWL (Web Ontology Language). Cette représentation formelle peut être utilisée par la suite comme un support à des opérations de raisonnement dans des contextes différents comme la recherche de documents sur le Web ou dans l'indexation sémantique des pages Web, etc.

2 Etat de l'art

D'après (Costa et al., 2005), une OP est « *an explicit, formal knowledge representation that expresses knowledge about a domain of application*». Les OPs permettent de décrire les connaissances sur un domaine d'une manière raisonnée, structurée et partageable, idéalement dans un format qui peut être lu et traité par un ordinateur et d'intégrer l'incertitude à ces

connaissances. Un RB est un modèle probabiliste qui est capable de gérer l'incertain. Il est défini par un graphe orienté sans circuit dont les sommets représentent des variables aléatoires d'un domaine, les arcs indiquent des dépendances conditionnelles entre les sommets et des probabilités conditionnelles permettent de quantifier les dépendances entre les nœuds. Il repose sur la théorie de graphe pour représenter les dépendances conditionnelles entre les variables d'un système étudié, et sur la théorie de probabilités pour définir mathématiquement ces dépendances (Philippe L., 2006). Divers chercheurs ont essayé de combiner les RBs avec les ontologies afin de représenter et raisonner avec l'incertitude : (Costa et al., 2005), (Yang Y. & Calmet J., 2005), (Ding & Peng, 2004) ont proposé des extensions de la formalisme standard OWL qui sont BayesOWL, OntoBayes et PR-OWL. (Koller D. & Levy A., 1997), (Fabio G. et al, 2011), etc ont proposé des nouvelles extensions probabilistes de logiques de description probabilistes (LDP) qui sont une famille de langages de représentation de connaissances ontologiques probabilistes et qui s'adoptent bien au Web.

Nous avons intégré le RB dans l'ontologie afin de remédier les défauts de chacune (le RB est incapable de représenter formellement les informations incertaines et l'ontologie est incapable de gérer l'incertitude) et de combiner les inférences bayésiennes (mise à jour de probabilités) avec l'inférence ontologique (vérification d'instances, consistance, etc). L'OP obtenue peut être utilisée par la suite dans des applications d'aide à la décision qui doivent être capable de raisonner avec des connaissances incertaines.

3 Notre méthode de construction d'OP

Un des principaux avantages de RB est leur capacité de gérer l'incertitude : cela permet de représenter les informations incertaines avec une manière très lisible et clair et offre des techniques puissants d'inférences bayésiennes qui consiste à propager une ou plusieurs informations (valeurs probabilistes) dans le réseau pour en déduire comment ceci intervient sur les probabilités d'autres variables. Cette inférence est appelée "mise à jour" des probabilités. Pour formaliser la représentation de ce modèle et la rendre compréhensible par les machines, nous proposons une méthode de modélisation formelle de RB en utilisant la formalisme standard d'ontologie OWL. L'objectif de cette méthode n'est pas seulement de représenter formellement les informations provenant de RB sous la forme d'une OP mais aussi de combiner quelques points forts de RB avec ceux de l'ontologie : les outils de raisonnement sur l'ontologie (subsomption, vérification d'instances, consistance, ...) avec les outils offerts par le RB (représentation des informations incertaines, l'inférence,…) afin de fournir des mécanismes permettant de représenter les informations incertaines et d'inférer des connaissances implicites à partir des connaissances explicites stockées dans une ontologie probabiliste.

Soit Q est un réseau bayésien qui est défini par un graphe G=(A, S), avec A= $\{A_1, A_2, \ldots, A_m\}$ est l'ensemble de relations de dépendance de taille m et S=$\{S_1, S_2, \ldots, S_n\}$ est l'ensemble de sommets de taille n. On observe la similitude de structure entre le graphe de RB et le graphe d'ontologie : Le RB et l'ontologie sont définis par un graphe. Ces deux graphes possèdent un ensemble d'arcs et un ensemble de sommets. Cette similitude facilite la tâche de transformation de RB en une OP qui peut être faite à l'aide d'un ensemble de règles de traduction qui sont :

1) Le n nœuds de graphe de RB sont transformés en un ensemble de concepts C= $\{c_1, c_2, \ldots, c_n\}$, avec n est le nombre de concepts d'ontologie.

2) Les valeurs possibles de chaque nœuds sont transformées en ensemble d'instances de concepts I. I=$\{I_1, I_2, \ldots, I_n\}$ et $I_j=\{i_1, i_2, \ldots, i_k\}$, avec I_j est l'ensemble d'instance de concept j et k est le nombre d'instances de concept j.

3) La probabilité à priori P(A) est converti en une valeur probabiliste (entre 0 et 1) d'une propriété de type DataProperty, nommée «ProbApriori-ci», de concept ci. $ci \in Rac$, avec $Rac = \{c1, c2, \ldots, cl\}$ est l'ensemble de racines de RB (Rac sous ensemble de

C), et l est le nombre de nœuds sans parents (racines). $ci \in Rac$ si et seulement si $ci \in C$ et $Parent(ci) = \emptyset$.

4) La présence d'un arc entre deux nœuds de RB présente une relation de dépendance entre ces deux nœuds. Cette relation de dépendance sera transformée en une relation N-aire (une relation N-aire est une relation ayant plus de deux arguments ou un attribut additionnel.), nommée «dépend-de-j», qui est caractérisée par une propriété de type DataProperty , nommée «ProbCond-j». Cette propriété exprime la probabilité conditionnelle P(A|B), qui indique la probabilité pour chaque valeur du A, étant donné les combinaisons de valeurs de parents du A (B). Donc, le m arcs seront convertis en un ensemble de relations N-aire entre les concepts R={R_1 : dépend-de-1, R_2 : dépend-de-2, …, R_m : dépend-de-m}. Chaque relation R_j de R est caractérisée par une propriété de type DataProperty : «ProbCond-j».

4 Conclusion

Dans cet article, nous avons proposé une méthode permettant de représenter sémantiquement un RB sous la forme d'une OP. Dans le futur travail, nous essayerons d'utiliser cette OP dans des applications réelles et nous nous intéresserons plus à l'inférence dans l'OP : Quels sont les axiomes (ajoutés dans OP) que nous pouvons déduire à partir de RB ? Quelles sont les nouvelles connaissances que nous pouvons extraire ou inférer à partir de cette ontologie probabiliste ?...

Références

Studer R. et al. (1998). Studer R. Benjamins, V.R., Fensel, D., Knowledge engineering: Principles and methods. Data & Knowledge Engineering 25, 161–197.

Nottelmann H. & Fuhr N. (2006). Adding probabilities and rules to OWL Lite subsets based on probabilistic Datalog, Int. J. Uncertain. Fuzz. pp 17–42.

Heinsohn J. (1994). Probabilistic description logics, in: Proceedings UAI-1994, Morgan Kaufmann, pp. 311–318.

Pool M. & Aikin J. (2004). KEEPER and Protégé: An elicitation environment for Bayesian inference tools, in: Proceedings of the Workshop on Protégé and Reasoning held at the IPC.

da Costa P.C.G. et al. (2005). da Costa P.C.G. et Laskey K.B. Laskey K.J., PR-OWL: A Bayesian ontology language for the Semantic Web, in: Proceedings URSW, pp. 23–33.

Philippe L. (2006). Réseaux bayésiens : apprentissage et modélisation de systèmes complexes, HABILITATION.

Yang Y. & Calmet J. (2005). Ontobayes: An ontology-driven uncertainty model. In: (IAWTIC'05). IEEE Computer Society (2005) 457-464.

Ding Z. & Peng Y. (2004). A probabilistic extension to ontology language OWL, in: Proceedings HICSS.

Koller D. & Levy A. (1997). P-CLASSIC: A tractable probabilistic description logic, in: Proceedings AAAI-1997, AAAI Press/MIT Press, pp. 390–397.

Fabio G. et al (2011). Rodrigo B. Polastroa, Fabio G. Cozmana, Felipe I. Takiyamaa, and Kate C. Revoredob, Computing Inferences for Credal ALC Terminologies.

Stanjanovic L. (2004). Methods and Tools for Ontology Evolution, Thèse de doctorat de l'Université de Karlsruhe.