Mesures de responsabilité à faible complexité pour requêtes en présence d'ontologies

Meghyn Bienvenu, Diego Figueira, Pierre Lafourcade

Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR5800, F-33400 Talence, France

{meghyn.bienvenu, diego.figueira, pierre.lafourcade}@u-bordeaux.fr

Résumé

Des travaux récents visant à expliquer des réponses à des requêtes ont mis au jour différentes mesures de responsabilité, qui attribuent un score à chaque fait d'une base de données, afin quantifier sa contribution à une réponse donnée. Dans cet article, nous étudions la complexité du calcul de telles mesures en présence d'ontologies. Nous nous concentrons sur une famille de mesures récemment introduites à partir de la valeur de Shapley, et qui s'expriment comme des sommes pondérées de supports minimaux (abrégé WSMS en anglais), dont nous explorons à la fois la complexité combinée et celle des données.

Mots-clés

Requêtes en présence d'ontologies, mesures de responsabilité, valeur de Shapley, analyse de complexité.

Abstract

Recent work on quantitative approaches to explaining query answers employs responsibility measures to assign scores to facts in order to quantify their respective contributions to obtaining a given answer. In this paper, we study the complexity of computing such responsibility scores in the setting of ontology-mediated query answering, focusing on a very recently introduced family of Shapley-value-based responsibility measures defined in terms of weighted sums of minimal supports (WSMS), exploring both their data and combined complexities for various ontology languages.

Keywords

Ontology-mediated query answering, responsibility measures, Shapley value, complexity analysis

☞ Cet article est un résumé de [3].

1 Introduction

Parmi les méthodes étudiées pour expliquer les réponses aux requêtes, les mesures quantitatives de responsabilité attribuent un score à chaque fait d'une base de données quantifiant sa contribution à la réponse d'une requête. Nous ne considèrerons ici que des requêtes booléennes et croissantes ; [1] il s'agira donc d'expliquer pourquoi la requête est satisfaite ($\mathcal{D} \models q$). Deux mesures notables ont été définies à partir de la valeur de Shapley Sh [5], méthode de distribution des richesses dans un jeu coopératif définie dans les années 50 : la valeur de Shapley dite *drastique* [4], et celle dite *MS*, plus récente [2]. Elles sont obtenues en modélisant la requête q et la base de données $\mathcal{D}$ par deux jeux différents : d'après le jeu $\xi^{\mathsf{dr}}_{q,\mathcal{D}}$ de la valeur dite drastique, une sous-base $S \subseteq \mathcal{D}$ reçoit 1 si $S \models q$ (et 0 sinon), tandis que pour le jeu $\xi^{\mathsf{ms}}_{q,\mathcal{D}}$ sous-jacent à la valeur MS, une sous-base S reçoit le nombre de *supports minimaux* (i.e. sous-ensembles minimaux M de faits tels que $M \models q$) présents dans S.

Dans le présent article, nous nous intéressons au calcul de la valeur de Shapley MS sur des données enrichies de connaissances au moyen d'une ontologie, à la suite de [1] qui avait fait de même pour la valeur de Shapley drastique. En effet, les résultats obtenus dans le cadre des bases de données (sans ontologie) ont montré que valeur de Shapley MS est généralement plus facile à calculer que la valeur drastique. De fait, $\mathrm{Sh}_{\xi^{\mathsf{ms}}_{q,\mathcal{D}}}(f)$ s'exprime plus simplement comme la somme, sur tous les supports minimaux M contenant f, de $\frac{1}{|M|}$. En outre, la valeur de Shapley MS se généralise à toute une famille de mesures (WSMS pour *weighted sums of minimal supports*) qui s'expriment facilement sous forme close, tout en restant chacune la valeurs de Shapley d'un certain jeu coopératif la caractérisant.

De plus, la valeur de Shapley MS (comme l'ensemble des mesures WSMS) est semblable à la valeur de Shapley drastique sur le plan conceptuel puisque toutes satisfont tous les desiderata suivants introduits dans [2, §4.1] et illustrés dans l'Exemple 1 : *(Sym-db)* deux faits indistinguables pour q doivent avoir la même valeur ; *(Null-db)* un fait, dit *pertinent*, qui appartient à un support minimal doit avoir une valeur strictement positive, et un fait non pertinent doit avoir une valeur nulle ; toutes choses égales par ailleurs, *(MS1)* un fait apparaissant dans des supports minimaux plus petits a une plus grande valeur, et *(MS2)* un fait apparaissant dans plus de supports minimaux a une plus grande valeur.

Exemple 1. *Considérons la base de connaissances* $(\mathcal{D}, \mathcal{O})$ *présentée en figure 1, ainsi que la requête booléenne* $q :=$ $\mathsf{AuPoisson}(soleCancalaise)$. *Il existe trois supports mini-*

1. Une requête est *booléenne* si elle est sans variables libres (une question fermée : «vrai» ou «faux»), et *croissante* si l'ajout de nouveaux faits préserve les réponses (ou la satisfaction). Des réponses aux requêtes avec variables libres sont traitées par réduction aux requêtes booléennes.

$\exists$Ingr.AuPoisson $\sqsubseteq$ AuPoisson   Garniture $\sqsubseteq$ Ingr
FruitMer $\sqsubseteq$ AuPoisson   Poisson $\sqsubseteq$ AuPoisson

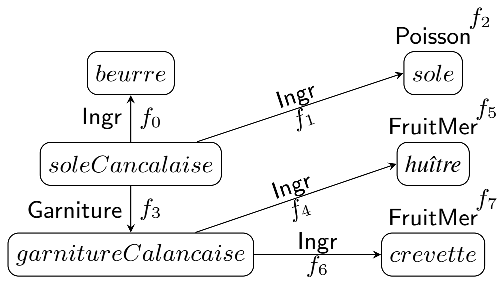

FIGURE 1 – Base de connaissances en logique de description comprenant des données $\mathcal{D}$ (bas) et une ontologie $\mathcal{O}$ (haut). Les flèches étiquetées indiquent des faits binaires et les étiquettes des nœuds (ex. Poisson) des faits unaires.

$\star$	f_0	f_1, f_2	f_3	f_4, f_5, f_6, f_7
dr	0	$\frac{1224}{5040}$	$\frac{1056}{5040}$	$\frac{384}{5040}$
ms	0	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$

TABLE 1 – Valeurs de $\mathrm{Sh}_{\xi^{\star}_{q,\mathcal{D}}}(\_)$ pour la figure 1. Les faits équivalents par (Sym-db) sont regroupées.

maux pour $Q := (\mathcal{O}, q)$ dans $\mathcal{D}$: $\{f_1, f_2\}$, $\{f_3, f_4, f_5\}$ et $\{f_3, f_6, f_7\}$. Les propriétés décrites ci-dessus s'appliquent comme suit : (Sym-db) donne $\phi_{Q,\mathcal{D}}(f_1) = \phi_{Q,\mathcal{D}}(f_2)$ (car f_1 et f_2 apparaissent dans les mêmes supports minimaux) ; (Null-db) donne $\phi_{Q,\mathcal{D}}(f_0) = 0$ (car f_0 n'est pas pertinent) ; (MS1) donne $\phi_{Q,\mathcal{D}}(f_1) > \phi_{Q,\mathcal{D}}(f_4)$ (car le support minimal où apparaît f_1 est plus petit) ; (MS2) donne $\phi_{Q,\mathcal{D}}(f_3) > \phi_{Q,\mathcal{D}}(f_4)$ (car f_3 apparaît dans deux supports minimaux contre un seul pour f_4). $\triangle$

2 Complexité

Complexité des données Etudions la complexité du problème $\mathsf{SVC}^{\mathsf{ms}}_Q$ de calculer $\mathrm{Sh}_{\xi^{\mathsf{ms}}_{Q,\mathcal{D}}}(f)$ pour un base de données $\mathcal{D}$ et un fait f en entrée, avec une requête en présence d'ontologie $Q = (\mathcal{O}, q)$ fixée. Comme nous l'avons vu, cette tâche revient à calculer une somme pondérée de supports minimaux, et donc à dénombrer les supports minimaux groupés par taille. Toute requête en présence d'ontologie pouvant se réécrire en *union de requêtes conjonctives* (UCQ) n'admettant que des supports minimaux de taille bornée (par la taille l de la réécriture), il est possible de les énumérer naïvement en temps polynomial ($\mathcal{D}^l$) pour les dénombrer et par suite en déduire $\mathsf{SVC}^{\mathsf{ms}}_Q$:

Theorem 2. $\mathsf{SVC}^{ms}_Q \in \mathsf{FP}$ *pour toute requête en présence d'ontologie Q admettant une réécriture en UCQ.*[2] *C'est en particulier le cas de tout $Q \in (\mathsf{DL\text{-}Lite}_{\mathcal{R}}, \mathsf{UCQ})$.*

2. Rappelons que toutes les requêtes considérées ici sont croissantes.

Pour le calcul pratique, cette énumération naïve peut, en réalité, être nettement améliorée puisque l'évaluation de $\mathsf{SVC}^{\mathsf{ms}}_q$ pour toute UCQ q peut s'effectuer par l'évaluation en parallèle d'un ensemble de requêtes SQL dérivées de q.

Si l'on s'écarte des requêtes admettant une réécriture en UCQ, en revanche, l'on trouve très vite des requêtes demandant intuitivement «existe-t-il un chemin entre a et b ?», i.e. des requêtes exprimant la notion d'accessibilité dans un graphe. Or les supports minimaux pour de telles requêtes sont les chemins simples, qui sont $\#$P-difficiles à compter.

Theorem 3. *Soit $\mathcal{L}$ une logique de description pouvant exprimer l'axiome $\exists r.A \sqsubseteq A$. Alors il existe une requête $(\mathcal{O}, q)$ avec $\mathcal{O} \in \mathcal{L}$ et q atomique tq SVC^{ms}_Q est $\#\mathsf{P}$-difficile.*

Complexité combinée Si nous avons établi que $\mathsf{SVC}^{\mathsf{ms}}_Q \in \mathsf{FP}$ pour tout Q se réécrivant en UCQ, la complexité combinée (qui ne suppose donc plus que $|Q|$ est constante) devient $\#$P-difficile à la moindre conjonction dans les axiomes :

Proposition 4. SVC^{ms}_Q *est $\#\mathsf{P}$-diff. pour la logique de description $\mathcal{L}_{\sqcap}$ autorisant les axiomes de la forme $A \sqcap B \sqsubseteq C$.*

Si, en revanche, l'on se restreint à des logiques comme $\mathsf{DL\text{-}Lite}_{\mathcal{R}}$ pour lesquelles les supports minimaux de requêtes atomiques sont toujours des singletons, il est possible de recouvrer certains des résultats positifs connus pour les requêtes sans onotologie. Plus précisément, l'on sait que $\mathsf{SVC}^{\mathsf{ms}}_{\mathcal{C}} \in \mathsf{FP}$ pour toute classe $\mathcal{C}$ de *requête conjonctive* (CQ) sans auto-jointure (i.e. sans atomes distincts du même prédicat) et de largeur d'arbre bornée. La notion d'auto-jointure n'est pas pertinente en présence d'ontologie, parce que des axiomes $\{A \sqsubseteq B, B \sqsubseteq A\}$ peuvent confondre les prédicats A et B, mais l'on peut capturer son esprit par la notion d'*interaction*, qui intervient dès qu'un fait unique peut satisfaire deux atomes distincts de la requête.[3]

Theorem 5. *Pour toute classe $\mathcal{C} \subseteq (\mathsf{DL\text{-}Lite}_{\mathcal{R}}, \mathsf{CQ})$ telle que tout $Q \in \mathcal{C}$ est sans interaction et que l'ensemble $\{q \mid (\mathcal{O}, q) \in \mathcal{C}\}$ ait une largeur d'arbe bornée,* $\mathsf{SVC}^{ms}_{\mathcal{C}} \in \mathsf{FP}$.

Références

[1] Meghyn Bienvenu, Diego Figueira, and Pierre Lafourcade. Shapley value computation in ontology-mediated query answering. In *KR*, 2024.

[2] Meghyn Bienvenu, Diego Figueira, and Pierre Lafourcade. Shapley revisited : Tractable responsibility measures for query answers. In *PODS*, 2025.

[3] Meghyn Bienvenu, Diego Figueira, and Pierre Lafourcade. Tractable responsibility measures for ontology-mediated query answering. In *KR*, 2025.

[4] Ester Livshits, Leopoldo Bertossi, Benny Kimelfeld, and Moshe Sebag. The Shapley value of tuples in query answering. *LMCS*, Volume 17, Issue 3 :6942, 2021.

[5] Lloyd Stowell Shapley. A value for n-person games. *Contributions to the Theory of Games*, Volume II :307–318, 1953.

3. La véritable notion d'interaction est plus subtile : il est notamment possible d'avoir une interaction avec un seul atome, mais l'idée reste celle d'un fait satisfaisant l'atome de deux « manières différentes ».