Gestion des connaissances pour l'aide à la décision/ Knowledge management for decision aid

Souhila KACI

Partie 3/Part 3
Fusion d'informations/Information merging

Fusion d'informations

Outline

- Fusion de bases propositionnelles sans priorité
- Fusion de bases propositionnelles avec priorité implicite
- Fusion de bases propositionnelles avec priorité explicite

Outline

- Fusion de bases propositionnelles sans priorité
- Fusion de bases propositionnelles avec priorité implicite
- Fusion de bases propositionnelles avec priorité explicite

Fusion de bases propositionnelles sans priorité

- Union des bases
- Le résultat de l'union est souvent un ensemble incohérent
- On applique les méthodes de gestion de l'incohérence vues précédemment (argumentation, SMC, etc)

Outline

- Fusion de bases propositionnelles sans priorité
- Fusion de bases propositionnelles avec priorité implicite
- Fusion de bases propositionnelles avec priorité explicite

Fusion de bases propositionnelles avec priorité implicite (1)

- Les informations sont codées en logique propositionnelle
- Aucune relation d'ordre n'est associée à ces informations
- Une relation d'ordre implicite entre les informations est calculée
- Des opérateurs d'agrégation sont définis sur la base de cette relation d'ordre

L'idée...

Mesurer la proximité des interprétations (outcomes) par rapport aux informations, en se basant pour cela sur le calcul de distances.

Fusion de bases propositionnelles avec priorité implicite (2)

Processus de fusion – Trois étapes :

1. Etape 1 : Calculer la proximité de chaque interprétation des bases à fusionner. On calcule une distance locale. Un pré-ordre total sur Ω est calculé par rapport à chacune des bases.
2. Etape 2 : Calculer un pré-ordre sur Ω par rapport à l'ensemble des bases à fusionner. Il est obtenu au moyen d'une distance globale; résultat de l'agrégation des distances locales.
3. Etape 3 : Calculer le résultat de la fusion.

Example 1

Un professeur demande à ses trois étudiants lesquels parmi les langages suivants *SQL* (noté s), O_2 (noté o) et *Datalog* (noté d) ils souhaiteraient étudier.

- Le premier étudiant désire étudier seulement *SQL* ou O_2 mais pas *Datalog* : $K_1 = (s \vee o) \wedge \neg d$.
- Le deuxième ne veut pas étudier *SQL* et préfère étudier soit O_2 soit *Datalog* mais pas les deux à la fois : $K_2 = (\neg s \wedge d \wedge \neg o) \vee (\neg s \wedge \neg d \wedge o)$.
- Le troisième veut étudier les trois langages : $K_3 = (s \wedge d \wedge o)$.

Soit $E = \{K_1, \cdots, K_n\}$ $(n \geq 1)$ un multi-ensemble de n bases propositionnelles cohérentes (non nécessairement différentes) à fusionner.

- La distance locale est la distance entre une interprétation ω et une base K_i. Elle est notée $d(\omega, K_i)$.
- Elle est basée sur la distance entre deux interprétations ω et ω', appellée distance de Hamming et notée $H(\omega, \omega')$. $H(\omega, \omega') =$ nombre de littéraux différents entre ω et ω'.
- $d(\omega, K_i) = \min_{\omega' \models K_i} H(\omega, \omega')$.

La distance locale permet de générer un pré-ordre sur Ω par rapport à chacune des bases à fusionner même si ces dernières ne sont pas stratifiées.

$$\forall \omega, \omega' \in \Omega, \omega \succeq_K \omega' \text{ ssi } d(\omega, K) \leq d(\omega', K).$$

→ À chaque fois, calculer les modèles (distance 0) et calculer la distance minimale entre chaque interprétation et un modèle (le plus proche donc)

Etape 1 : Calcul des distances locales (2)

Intuitivement, si l'agent exprime des préférences alors ces dernières sont vues comme un ensemble de sous-buts indépendants à satisfaire. Une interprétation est considérée comme totalement satisfaisante si elle satisfait tous les sous-buts, et entre deux interprétations qui ne satisfont pas tous les sous-buts, l'une est considérée plus satisfaisante que l'autre si elle satisfait plus de sous-buts.

Etape 1 : Calcul des distances locales (3)

Considérons le premier étudiant (étudier seulement *SQL* ou O_2 mais pas *Datalog*).

Le professeur suppose que l'étudiant exprime deux buts indépendants :

- vouloir apprendre *SQL*, ou O_2, ou les deux ($s \vee o$),
- et ne pas vouloir apprendre *Datalog* ($\neg d$).

A partir de ces buts, on ordonne les différentes situations possibles comme suit :

1. les situations $s\neg do$, $\neg s\neg do$ et $s\neg d\neg o$ sont préférées puisqu'elles satisfont les deux sous-buts $\neg d$ et $s \vee o$,
2. les situations $\neg s\neg d\neg o$, sdo, $sd\neg o$ et $\neg sdo$ sont moins préférées que les précédentes car elles satisfont un seul sous-but,
3. et enfin la situation $\neg sd\neg o$ est la moins préférée puisqu'elle falsifie les deux sous-buts.

Etape 1 : Calcul des distances locales (4)
Exemple 1 (suite)

$K_1 : (s \vee o) \wedge \neg d$

$K_2 : (\neg s \wedge d \wedge \neg o) \vee (\neg s \wedge \neg d \wedge o)$

ω	$d(\omega, K_1)$	$d(\omega, K_2)$	$d(\omega, K_3)$
$\omega_0 : \neg s \neg d \neg o$	1	1	3
$\omega_1 : \neg s \neg d o$	0	0	2
$\omega_2 : \neg s d \neg o$	2	0	2
$\omega_3 : \neg s d o$	1	1	1
$\omega_4 : s \neg d \neg o$	0	2	2
$\omega_5 : s \neg d o$	0	1	1
$\omega_6 : s d \neg o$	1	1	1
$\omega_7 : s d o$	1	2	0

Etape 2 : Calcul des distances globales (1)

Définir une fonction qui, à partir des distances locales, calcule une distance *globale* obtenue par l'agrégation des distances locales avec un opérateur d'agrégation, noté $\oplus$. Cette distance est notée $d_{\oplus}(\omega, E)$.

Bases d'inégales importances

Quand les bases n'ont pas la même importance, une affectation de poids est utilisée qui consiste à attribuer des poids aux bases, définissant ainsi leur niveau d'importance.

$$d_{\mathcal{WS}}(\omega, E) = \sum_{i=1}^{n} k_i * d(\omega, K_i),$$

où les k_i $(i = 1, \cdots, n)$ sont des entiers positifs associés aux bases K_i représentant leur niveau d'importance. Ce n'est pas la somme pondérée habituelle où $\sum_{i=1}^{n} k_i = 1$.

$$\forall \omega, \omega' \in \Omega,\ \omega \preceq_E^{\mathcal{WS}} \omega' \text{ ssi } d_{\mathcal{WS}}(\omega, E) \leq d_{\mathcal{WS}}(\omega', E).$$

Tendance majoritaire

Satisfaire la *majorité* des bases.

$$d_{\sum}(\omega, E) = \sum_{i=1}^{n} d(\omega, K_i).$$

$$\forall \omega, \omega' \in \Omega,\ \omega \succeq_E^{\sum} \omega' \text{ ssi } d_{\sum}(\omega, E) \leq d_{\sum}(\omega', E).$$

Opérateur égalitariste idempotent

Satisfaire *toutes* les bases.

$$d_{\mathcal{MAX}}(\omega, E) = \max_{i=1,\cdots,n} d(\omega, K_i).$$

$$\forall \omega, \omega' \in \Omega,\ \omega \succeq_E^{\mathcal{MAX}} \omega' \text{ ssi } d_{\mathcal{MAX}}(\omega, E) \leq d_{\mathcal{MAX}}(\omega', E).$$

Etape 2 : Calcul des distances globales (3)

Opérateur égalitariste basé sur le Lexi-max (ou $\mathcal{MAX}$ généralisé)

Associer à chaque interprétation un vecteur ordonné (ordre décroissant) des distances locales. Ce vecteur est noté $d_{\mathcal{GMAX}}(\omega, E)$. Ensuite, on applique l'ordre lexicographique sur les vecteurs associés aux interprétations.

$$\forall \omega, \omega' \in \Omega,\ \omega \succeq_E^{\mathcal{GMAX}} \omega' \text{ ssi } d_{\mathcal{GMAX}}(\omega, E) \leq d_{\mathcal{GMAX}}(\omega', E).$$

Opérateur prudent idempotent

$$d_{\mathcal{MIN}}(\omega, E) = \min_{i=1,\cdots,n} d(\omega, K_i).$$

$$\forall \omega, \omega' \in \Omega,\ \omega \succeq_E^{\mathcal{MIN}} \omega' \text{ ssi } d_{\mathcal{MIN}}(\omega, E) \leq d_{\mathcal{MIN}}(\omega', E).$$

Etape 3 : Calcul du résultat de la fusion

Le but de l'étape précédente était de calculer un pré-ordre $\succeq_E^{\oplus}$ sur Ω globalement par rapport à toutes les bases. Les interprétations préférées sont celles qui se rapprochent le plus de l'ensemble de toutes les bases. Ces interprétations sont les préférées par rapport à $\succeq_E^{\oplus}$. Le résultat de la fusion, noté $Merge_{\oplus}(E)$, est l'ensemble de ces interprétations.

$$Merge_{\oplus}(E) = \max(\Omega, \succeq_E^{\oplus}).$$

Exemple (suite)

$k_1 = 1,\ k_2 = 3,\ k_3 = 1$ pour $\oplus = \mathcal{WS}$.

ω	d_1	d_2	d_3	$\sum$	$\mathcal{WS}$	$\mathcal{MAX}$	$\mathcal{MIN}$	$\mathcal{GMAX}$
ω_0	1	1	3	5	7	3	1	(3,1,1)
ω_1	0	0	2	2	2	2	0	(2,0,0)
ω_2	2	0	2	4	4	2	0	(2,2,0)
ω_3	1	1	1	3	5	1	1	(1,1,1)
ω_4	0	2	2	4	8	2	0	(2,2,0)
ω_5	0	1	1	2	4	1	0	(1,1,0)
ω_6	1	1	1	3	5	1	1	(1,1,1)
ω_7	1	2	0	3	7	2	0	(2,1,0)

Analyse de la fusion à base de la distance de Hamming

Problème

L'associativité n'est pas garantie même si l'opérateur d'agrégation est associatif.
Dans notre exemple, $Merge_{\sum}(\{K_1, K_2, K_3\}) \neq Merge_{\sum}(\{K', K_3\})$, avec $K' = Merge_{\sum}(\{K_1, K_2\})$.

ω	$d(\omega, K_1)$	$d(\omega, K_2)$	$d_{\sum}(\omega, \{K_1, K_2\})$
$\omega_0 : \neg s \neg d \neg o$	1	1	2
$\omega_1 : \neg s \neg d o$	0	0	0
$\omega_2 : \neg s d \neg o$	2	0	2
$\omega_3 : \neg s d o$	1	1	2
$\omega_4 : s \neg d \neg o$	0	2	2
$\omega_5 : s \neg d o$	0	1	1
$\omega_6 : s d \neg o$	1	1	2
$\omega_7 : s d o$	1	2	3

Distance locale par rapport à K'

ω	$d(\omega, K')$
$\omega_0 : \neg s \neg d \neg o$	1
$\omega_1 : \neg s \neg d o$	0
$\omega_2 : \neg s d \neg o$	2
$\omega_3 : \neg s d o$	1
$\omega_4 : s \neg d \neg o$	2
$\omega_5 : s \neg d o$	1
$\omega_6 : s d \neg o$	3
$\omega_7 : s d o$	2

Fusion de K' et K_3

ω	$d(\omega, K')$	$d(\omega, K_3)$	$d_{\sum}(\omega, \{K', K_3\})$
$\omega_0 : \neg s \neg d \neg o$	1	3	4
$\omega_1 : \neg s \neg d o$	0	2	2
$\omega_2 : \neg s d \neg o$	2	2	4
$\omega_3 : \neg s d o$	1	1	2
$\omega_4 : s \neg d \neg o$	2	2	4
$\omega_5 : s \neg d o$	1	1	2
$\omega_6 : s d \neg o$	3	1	4
$\omega_7 : s d o$	2	0	2

Outline

- Fusion de bases propositionnelles sans priorité
- Fusion de bases propositionnelles avec priorité implicite
- Fusion de bases propositionnelles avec priorité explicite

Fusion en logique possibiliste

- Input : $\mathcal{B} = \{B_1, \cdots, B_n\}$, un opérateur d'agrégation $\oplus$
- Output : $\mathcal{B}_{\oplus} = \{(D_j, 1 - \oplus(x_1, \cdots, x_n)) : j = 1, \cdots, n\}$, où D_j sont les disjonctions de taille j entre les formules ϕ_i prises des différentes bases B_i ($i = 1, \cdots, n$) et x_i est égal à $1 - a_i$ si $\phi_i \in D_j$. Il est égal à 1 sinon.

Cas de fusion de deux bases $B_1 = \{(\phi_i, a_i)\}$ et $B_2 = \{(\psi_j, b_j)\}$

Le résultat de la fusion est composé

- des bases initiales, avec cependant de nouveaux poids :
$$\{(\phi_i, 1 - \oplus(1 - a_i, 1)) : (\phi_i, a_i) \in B_1\} \cup$$
$$\{(\psi_j, 1 - \oplus(1, 1 - b_j)) : (\psi_j, b_j) \in B_2\},$$
- et des informations communes à B_1 et B_2 définies par :
$\{(\phi_i \vee \psi_j, 1 - \oplus(1 - a_i, 1 - b_j)) : (\phi_i, a_i) \in B_1, (\psi_j, b_j) \in B_2\}$.

Exemples de fusion en logique possibiliste

- $\mathcal{B}_{\min} = B_1 \cup B_2$
- $\mathcal{B}_{\max} = \{(\phi_i \vee \psi_j, \min(a_i, b_j)) | (\phi_i, a_i) \in B_1, (\psi_j, b_j) \in B_2\}$
- $\mathcal{B}_* = B_1 \cup B_2 \cup \{(\phi_i \vee \psi_j, a_i + b_j - a_i * b_j) : (\phi_i, a_i) \in B_1, (\psi_j, b_j) \in B_2\}$

Le résultat de la fusion est associatif lorsque l'opérateur d'agrégation est associatif.

Propriétés des opérateurs de fusion

Opérateur min

- Le résultat de la fusion peut être incohérent même si chacune des bases à fusionner est cohérente.
- Pas de renforcement des informations redondantes.

Opérateur max

- Le résultat de la fusion est cohérent dès qu'une des bases à fusionner est cohérente.
- Pas de renforcement des informations redondantes.

Opérateurs *

- Le résultat de la fusion peut être incohérent même si chacune des bases à fusionner est cohérente.
- Renforcement des informations redondantes.

Applications de la fusion d'informations

- Robotique
- Systèmes multi-agents
- Décision collective : Fusion de préférences, Agrégation des jugements

Exercice (1)

- Soient c, s et e trois variables propositionnelles qui représentent respectivement "le chat est dans la salle", "la souris est dans la salle", "la souris est effrayée".
- Soient B_1 et B_2 deux bases possibilistes décrivant les observations de deux agents d'un site composé d'une salle, d'un chat et d'une souris :
  - $B_1 = \{(\neg c \vee \neg s \vee e, 1), (\neg c, .8), (s, .5)\}$
  - $B_2 = \{(\neg c \vee \neg s \vee e, 1), (\neg c, .6), (e, .2)\}$

Questions

1. Interpréter les observations de chaque agent.
2. Fusionner les deux bases avec les opérateurs vus précédemment.
3. Quelles sont vos observations ?

Exercice (2)

- $B_1 = \{(\neg c \vee \neg s \vee e, 1), (\neg c, .8), (s, .5)\}$
- $B_2 = \{(\neg c \vee \neg s \vee e, 1), (\neg c, .6), (e, .2)\}$

- $\mathcal{B}_{min} = B_1 \cup B_2 = \{(\neg c \vee \neg s \vee e, 1), (\neg c, .8), (s, .5), (e, .2)\}$
- $\mathcal{B}_{max} = \{(\neg c \vee \neg s \vee e, 1), (\neg c, .6), (s \vee e, .2)\}$
- $\mathcal{B}_{*} = B_1 \cup B_2 \cup \{(\neg c \vee \neg s \vee e, 1), (\neg c, .92), (s \vee e, .6)\}$
  $= \{(\neg c \vee \neg s \vee e, 1), (\neg c, .92), (s \vee e, .6), (s, .5), (e, .2)\}$