

Révision d'une base d'informations personnelles avec gestion du vieillissement des informations

Salma Chaieb, Véronique Delcroix,
Ali Ben Mrad, Brahim Hnich, Emmanuelle Grislin

LAMIH UMR 8201
Université de Valenciennes et du Hainaut-Cambrésis
qui devient
Université Polytechnique des Hauts de France (UPHF)

JFRB - Toulouse, 30 mai - 1er juin 2018

Plan

- 1 Le contexte
 - Prévention des chutes des personnes âgées
- 2 Problématique et état de l'art
 - Problématique du vieillissement des informations
 - Comparaison avec la Théorie de la révision des connaissances
- 3 Proposition pour la révision d'une base d'informations
 - Deux composants pour la révision d'une base d'informations
 - Principe et algorithme pour la révision d'une base d'informations
- 4 Conclusion

Plan

- 1 Le contexte
 - Prévention des chutes des personnes âgées
- 2 Problématique et état de l'art
 - Problématique du vieillissement des informations
 - Comparaison avec la Théorie de la révision des connaissances
- 3 Proposition pour la révision d'une base d'informations
 - Deux composants pour la révision d'une base d'informations
 - Principe et algorithme pour la révision d'une base d'informations
- 4 Conclusion

Les enjeux de la chute des PA

Vieillesse \Rightarrow augmentation de la fréquence des chutes

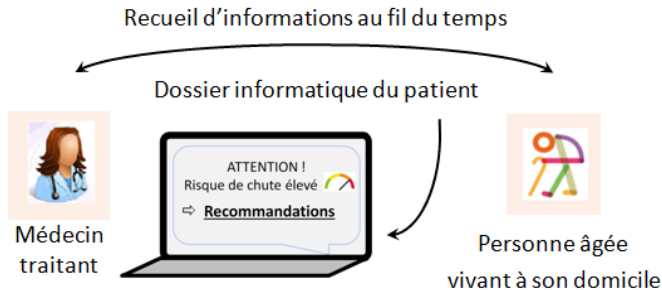
Chutes \Rightarrow blessures, hospitalisations, perte d'autonomie, entrée en institution mortalité, surtout chez les plus de 65 ans

Coût de la chute en France : évalué à 2 milliards d'euros par an.

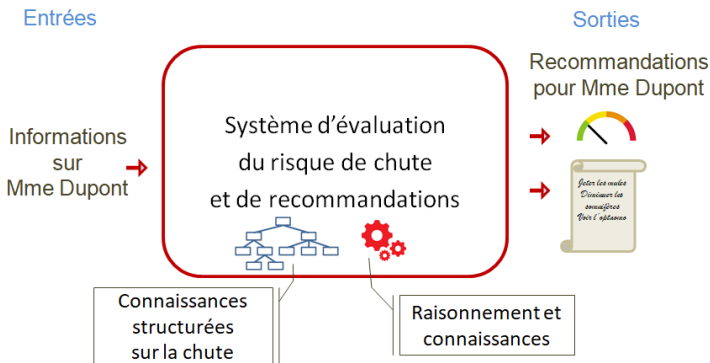
- 70 000 fractures de la hanche par an en France,
- 20% de ces chuteurs décèdent dans l'année qui suit l'opération.

Prévention des chutes : enjeu majeur pour la mobilité et la santé

Système de prévention des chutes des personnes âgées à destination des médecins traitants



Système de prévention des chutes des personnes âgées à destination des médecins traitants



Les défis de la prévention des chutes

Au quotidien, sur le terrain, Pas de temps pour ce problème

→ **Difficulté pour collecter les informations (qualité et quantité)**

Système de Gestion des informations personnelles (SGIP)

- Chargé de la collecte et de la gestion des informations d'une PA
- recueil des informations au fil du temps
- Objectif : fournir rapidement les informations sur la personne âgée demandées par un autre système
- Contient la fiche de la personne : liste d'observations

Contenu de la base d'informations personnelles ?

- Exemples
 - Morphologie, physiologie, informations médicales, géographiques, environnementales, familiales, comportement, capacités (et incapacités), compétences, goûts, etc.
- Liste guidée par les objectifs
- Cadre de la prévention de la chute
 - Facteurs de risques de chute et variables qui les influent
 - Extraction des connaissances, ontologie des facteurs de risques de chute des personnes âgées.

Plan

- 1 Le contexte
 - Prévention des chutes des personnes âgées
- 2 **Problématique et état de l'art**
 - Problématique du vieillissement des informations
 - Comparaison avec la Théorie de la révision des connaissances
- 3 Proposition pour la révision d'une base d'informations
 - Deux composants pour la révision d'une base d'informations
 - Principe et algorithme pour la révision d'une base d'informations
- 4 Conclusion

Une base d'informations personnelle (vision 1)

Information :

- Observation de certaines caractéristiques d'une personne faite à un moment → valeur d'une variable
- Valeur connue avec certitude au moment de l'observation.
- Comment savoir si une information est toujours vraie un certain temps plus tard ?

Variable	Valeur observée
laPersConduit	auMoins1ParSem
faitSesCourses	oui
dispositifGPS	ok

→ stocker la date d'observation

→ Durée de validité des informations ?

Problème lié au vieillissement des informations

Durée de validité d'une information :

Variable 1 : Sexe d'une personne : information à durée illimitée dans la très grande majorité des cas

Variable 2 : Capacité de marche : évolue avec l'âge (incertitude)

Problématique

Collecte les informations au fil du temps

+ vieillissement des informations

→ besoin d'évaluer la confiance associée à chaque information

→ Besoin de détecter et supprimer les informations obsolètes :

révision de la base d'informations

Théorie de la révision des connaissances

- Largement traité dans le cadre de la logique (Papini, 2000 ; Benferhat et al., 2017)
- Modèle logique des connaissances
- postulats AGM (Alchourrn, Grdenfors et Makinson, 1985) : liste de propriétés attendues d'un opérateur de révision

Révision des connaissances :

- modification de l'état épistémique d'un agent intelligent pour intégrer de nouvelles informations
- On révisé la base d'informations si elle est devenue incohérente suite à l'ajout de la nouvelle information
- On supprime des informations (le moins possible) pour maintenir la cohérence de la base de connaissances.

Comparaison (1)

	Révision des croyances / connaissances	Révision d'une base d'informations
Modèle de connaissance	logique	probabiliste
contexte de la révision	<i>monde</i> statique	informations anciennes susceptibles d'évoluer avec le temps

Comparaison (2)

	Révision des croyances / connaissances	Révision d'une base d'informations
Connaissance	susceptible d'être révisée	générale (population), non remise en cause
Information	même niveau que les connaissances	ponctuelle, locale, avec vieillissement
Représentation d'une information	formule propositionnelle	Observation datée d'une variable (déjà observée ou non)
Croyance	statut épistémique d'une formule (vrai, faux ou indéterminé)	état cognitif concernant une variable (observée ou pas) dans un contexte

Comparaison (3)

	Révision des croyances / connaissances	Révision d'une base d'informations
Révision	changement du statut épistémique des éléments identifiés (vrai \leftrightarrow faux <i>mais pas</i> \leftrightarrow <i>statut indéterminé</i>)	suppression des informations obsolètes identifiées
Contradiction	incompatibilité logique	probabilité <i>trop faible</i> de l'ensemble des observations
Critères de révision	minimalité de l'ensemble des éléments révisés, ordre partiel ou priorité, etc.	degré de confiance lié à l'âge des informations, probabilité de l'ensemble d'observations révisé

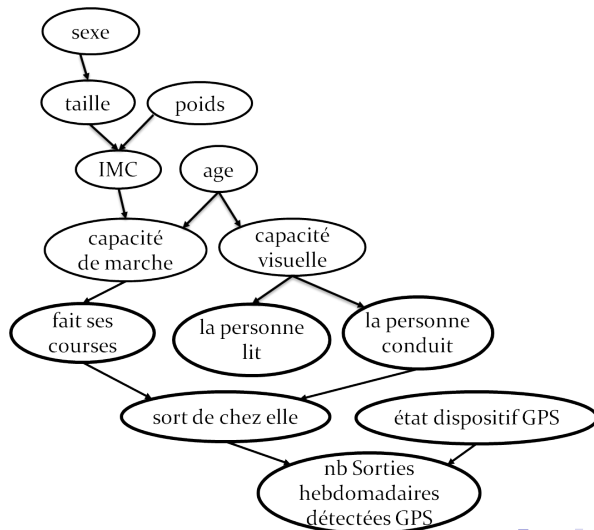
Plan

- 1 Le contexte
 - Prévention des chutes des personnes âgées
- 2 Problématique et état de l'art
 - Problématique du vieillissement des informations
 - Comparaison avec la Théorie de la révision des connaissances
- 3 Proposition pour la révision d'une base d'informations
 - Deux composants pour la révision d'une base d'informations
 - Principe et algorithme pour la révision d'une base d'informations
- 4 Conclusion

Deux composants pour la révision d'une base d'informations

- un réseau bayésien (RB) : embarque les connaissances générales sur les variables d'intérêt
- Une base d'informations : liste des variables du RB
 - si observée : valeur obs. + date obs. + degré de confiance lié à l'âge de l'obs.
 - si pas observée : croyance : $P(V | OBS)$

Un RB jouet



Base d'informations (BI)

Variable	Valeur observée	Date	Confiance	Le plus prob. obs	Prob. obs
laPersConduit	auMoins1ParSem	Date 1 : t_1	$Pm_1(t_1) = +$	-	-
faitSesCourses	oui	Date 1 : t_1	$Pm_1(t_1) = +$	-	-
dispositifGPS	ok	Date 1 : t_1	$Pm_1(t_1) = +$	-	-
capaVisuelle	-	-	-	correcte	0.98
sortDeChezElle	-	-	-	auMoins1ParSem	0.9
nbSortiesGPS	-	-	-	deuxOuPlus	0.73
capaMarche	-	-	-	normale	0.8
laPersLit	-	-	-	régulièrement	0.63
age	-	-	-	[60-63]	0.51
sexe	-	-	-	M	0.51
IMC	-	-	-	normale	0.38
taille	-	-	-	[160-170]	0.3
poids	-	-	-	[60-75]	0.19

L'ensemble des variables de la BI est fixe.

La BI contient une indication sur les variables observées *et* sur les variables non observées.

Principe de révision d'une base d'informations

Chaque fois qu'une nouvelle observation arrive dans la base d'information :

- 1 Est-ce qu'elle remet en cause les informations présentes dans la base ?
→ Détection des contradictions entre les nouvelles informations et les anciennes
- 2 Si oui, identification des observations obsolètes

Contradiction entre anciennes et nouvelles observations

Contradiction

Ensemble d'observations composés d'informations nouvelles et anciennes dont la probabilité jointe est nulle ou *tellement faible* que cela remet en cause la validité des observations anciennes.

Il y a contradiction si la probabilité jointe de toutes les observations est inférieure à un seuil ; (Ce seuil dépend de l'ensemble des variables observées).

base d'informations cohérente

Une base d'informations est cohérente si elle ne contient pas de contradiction.

Identification des informations obsolètes

Révision de la base d'informations : suppression des observations obsolètes de façon à lever les contradictions

Comment identifier les informations obsolètes ?

- par leur âge
 - Plus une information est ancienne, plus il est probable qu'elle soit devenue obsolète.
 - Variabilité entre les variables et pour une même variable : toutes les informations ne vieillissent pas de façon identique.
- par le raisonnement pour évaluer la vraisemblance des anciennes observations sachant les nouvelles

Degré de confiance d'une information

Degré de confiance associé à l'âge des informations

- fonction d'obsolescence
- dépend de la variable et de l'âge de l'information

Algorithme naïf de *révision* d'une base d'informations

- basé uniquement sur les degrés de confiance des informations
- une ancienne information est supprimée de la base dès que son degré de confiance est en dessous d'un seuil
- n'assure pas la *cohérence* de la base d'information
- compromis sur le seuil :
 - trop bas : risque élevé d'incohérence
 - trop haut : perte d'informations non obsolètes

Algorithme proposé pour la révision d'une base d'informations (BI)

- Input :
 - $\mathcal{O} = \{(X, x, d)\}$: les informations d'une BI cohérente
 - $\mathcal{Y} = (Y, y, d)$: Nouvelle information
 - RB sur toutes les variables de la BI
- Output : BI révisée cohérente
- Variables locales
 - \mathcal{O}_{old} : \mathcal{O} dont on enlève l'ancienne observations sur Y s'il elle existait
 - \mathcal{O}' : ens. d'informations à tester
 - $\mathcal{C}_{cand} \subset \mathcal{P}(\mathcal{O}_{old})$: des sous-ensembles d'observations candidats à expliquer la contradiction avec \mathcal{Y}

Algorithme pour la révision d'une base d'informations

Les éléments de l'ensemble \mathcal{C}_{cand} sont les observations susceptibles d'être obsolètes.

Cet ensemble est trié :

- degré de confiance décroissant
- singletons, puis paire, etc.

Algorithme proposé pour la révision d'une base d'informations (BI)

$$\mathcal{O}' = \mathcal{O}_{old} \cup \{\mathcal{Y}\}$$

Tant que \mathcal{O}' n'est pas cohérent

choisir $\mathcal{C} \in \mathcal{C}_{cand}$

$$\mathcal{O}' = \mathcal{O}_{old} \setminus \mathcal{C} \cup \{\mathcal{Y}\}$$

Fin TantQue

Return \mathcal{O}'

Comparaison

Avantages de l'algorithme proposé :

- assure la cohérence de la base d'informations
- ne supprime pas de la base des informations anciennes si elles ne génèrent pas de contradictions
- peut supprimer des informations peu anciennes si elles génèrent une contradiction avec la nouvelle information

Inconvénients

- la base d'informations révisée peut contenir des informations obsolètes qui ne génèrent aucune contradiction.
- → Possibilité de combiner les deux algorithmes en mettant un seuil de confiance très bas

Plan

- 1 Le contexte
 - Prévention des chutes des personnes âgées
- 2 Problématique et état de l'art
 - Problématique du vieillissement des informations
 - Comparaison avec la Théorie de la révision des connaissances
- 3 Proposition pour la révision d'une base d'informations
 - Deux composants pour la révision d'une base d'informations
 - Principe et algorithme pour la révision d'une base d'informations
- 4 Conclusion

Conclusion

Gestion du vieillissement des informations dans une base d'informations personnelles

- Problématique nouvelle, non traitée dans la littérature
- Intérêt confirmé par les médecins
- Difficulté de construire les modèles de connaissances
 - sélection des variables pertinentes sur la personne âgée (travail en cours) : facteurs de risque de chute et ce qui les influence
 - sur la façon dont vieillissent ces informations (aucun travail actuel)

Travail en cours

- Thèse de Salma Chaieb (Tunisie) : Gestion du vieillissement des informations dans une base d'informations personnelles
- Extraction et modélisation des connaissances sur les facteurs de risques de chute des personnes âgées

Questions ?