

Découverte de règles contextuelles pour prédire la présence d'amiante dans les bâtiments

IC 2021

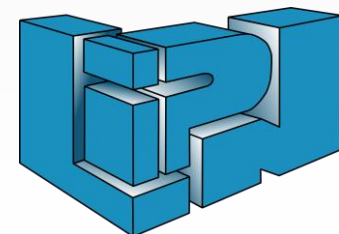
Thamer Mecharnia, Nathalie Pernelle, Celine Rouveirol , Fayçal Hamdi, Lydia Chibout Khelifa

CSTB
le futur en construction

LISN
LABORATOIRE INTERDISCIPLINAIRE
DES SCIENCES DU NUMÉRIQUE

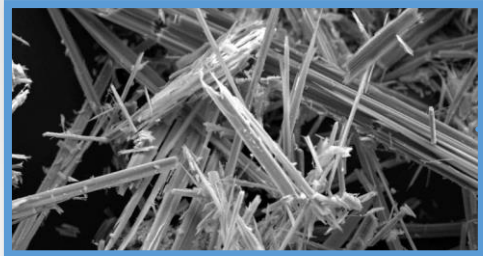
université
PARIS-SACLAY

Cédric
le cnam



Contexte

Amiante



Construction



- Asbestose ou fibrose pulmonaire
- Le cancer des poumons ou le mésothéliome



1997

Contexte

Documents CSTB



Diagnostic
amiante

Les composants du bâtiment et leurs relations partie-tout: structure, localisation et produit

Données temporelles :
année de construction

Diagnostiques de présence/absence
d'amiante dans un produit

Problématique

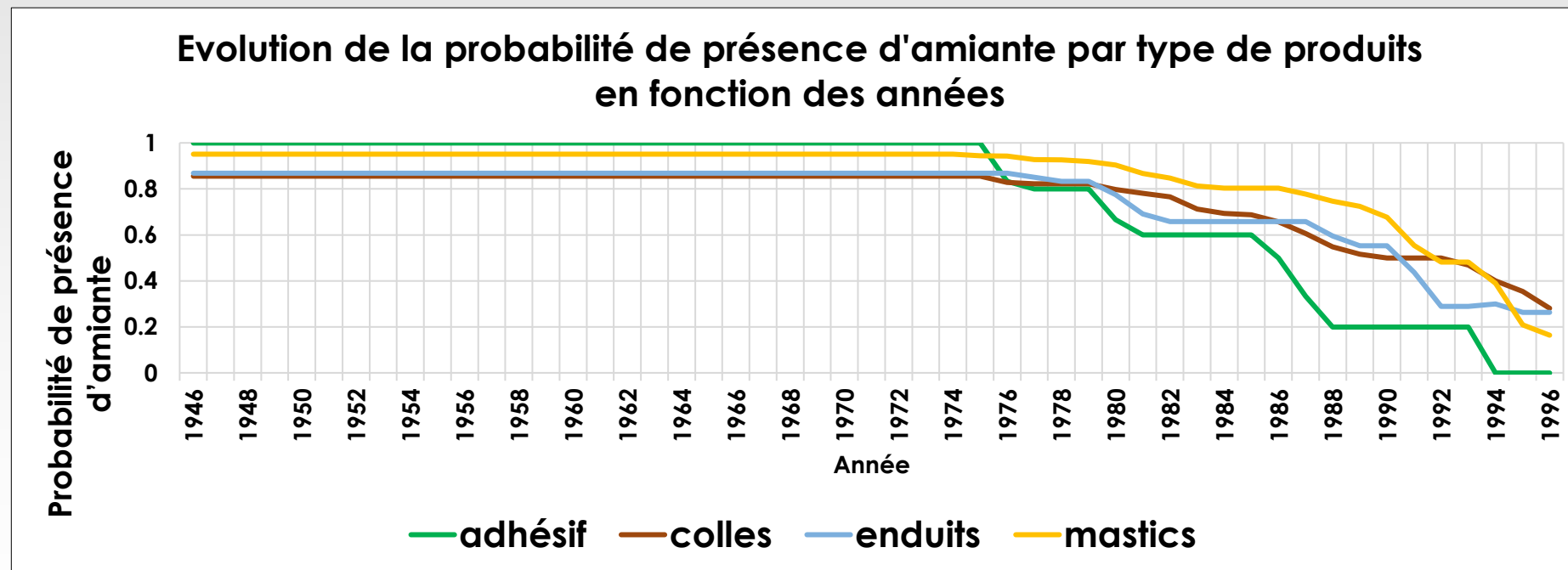
Information incomplète

Structure	Localisations	Produit	Diagnostic
PAROIS VERTICALES INTERIEURES	Murs	Enduit	0

Il existe de nombreux produits commercialisés de caractéristiques différentes de la classe de produits « **Enduit** »

Problématique

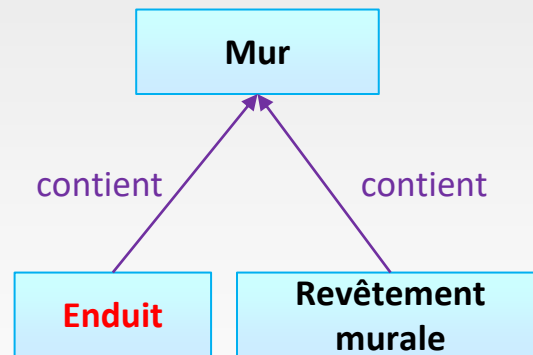
Influence des données temporelles



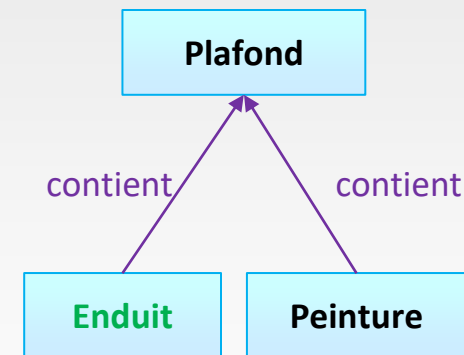
Construit à partir de l'INRS et l'ANDEVA

Problématique

Influence du contexte



Amianté



Non amianté

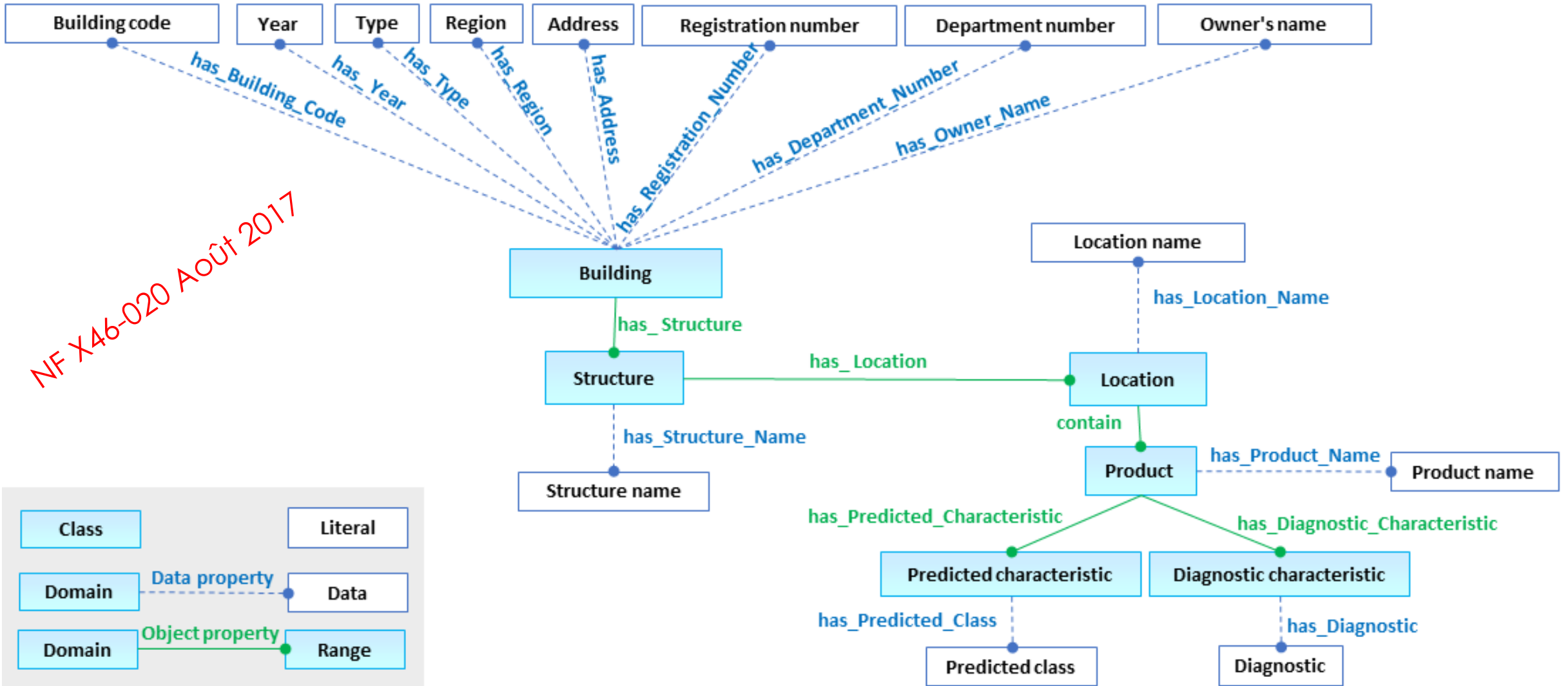
Comment classifier les produits comme amiantés ou non amiantés en exploitant le contexte et les données temporelles et incomplètes ?

Comment générer des règles de classification interprétables par l'expert ?

Sommaire

- L'ontologie Amiante
- Règles contextuelles
- CRA-Miner
- Expérimentations

L'ontologie Amiante



Règles contextuelles

Générées pour conclure sur la présence ou l'absence d'amiante dans un produit

Contexte conceptuel et Règles contextuelles

- Contexte conceptuel CO : ensemble de classes et de propriétés utilisables dans le corps de la règle.
- Règle Contextuelle : conjonction de prédicats qui conclut sur la présence ou l'absence d'amiante dans un produit :

$\vec{B} \rightarrow h$ avec $\vec{B} = \{B_1, B_2, \dots, B_n\}$ telle que $\forall B_i \in \vec{B}, \exists B_j \in CO \cup \{SWRL: CompareTO\}$ s. t. $B_i \sqsubseteq B_j$

h est le prédicat « has_diagnostic » qui est instancié par la valeur « positive » ou « négative ».

- Exemple de règle :

$colle(P), contain(L, P), has_location(S, L), has_structure(B, S), has_region(B, "Paris"), has_diagnostic_characteristic(P, D) \Rightarrow has_diagnostic(D, "positive")$

Mesures de qualité d'une règle contextuelle

- Support: nombre de prédictions correctes générées par la règle.
- Head coverage (hc): ratio entre le support et le nombre de diagnostics de type *val*

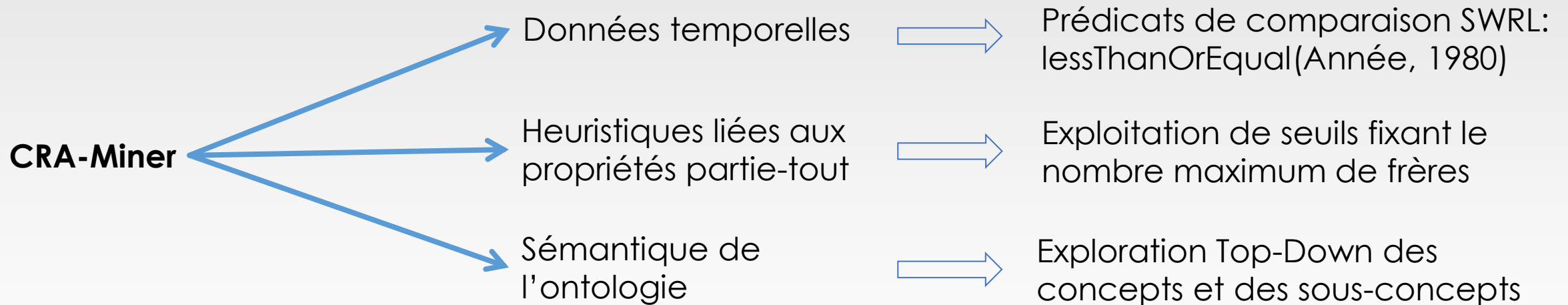
$$hc(\vec{B} \rightarrow has_diagnostic(D, val)) = \frac{supp(\vec{B} \rightarrow has_diagnostic(D, val))}{\#(D, val): has_diagnostic(D, val)}$$

- Confiance (conf) : ratio entre le support de la règle et le nombre de diagnostics différents qui participent à une instantiation du corps de la règle

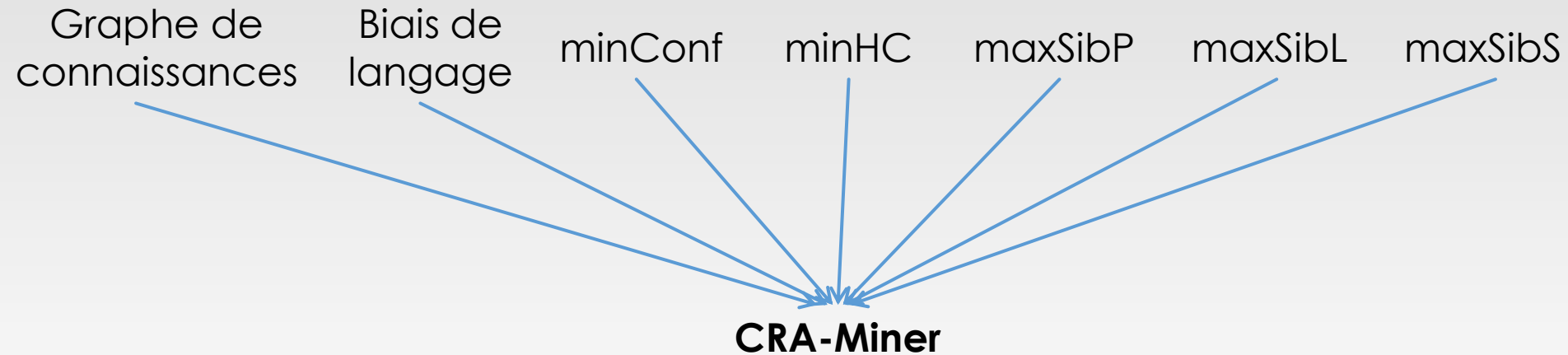
$$conf(\vec{B} \rightarrow has_diagnostic(D, val)) = \frac{supp(\vec{B} \rightarrow has_diagnostic(D, val))}{\#D: \exists x_1, \dots, x_n: \vec{B}}$$

L'objectif est de découvrir toutes les règles qui sont conformes aux contraintes du biais de langage et qui ont $hc \geq minHC$ et $conf \geq minConf$.

CRA-Miner



CRA-Miner



Toutes les règles contextuelles permettant de prédire la présence/absence d'amiante dans les produits

Etapes de CRA-Miner

CO = { product, location, structure, contain, has_location, has_region, has_year, has_structure, has_diagnostic_characteristic }

1. Spécialisation de T ($product(P), has_diagnostic_characteristic(P, D) \Rightarrow has_diagnostic(D, Value)$) en utilisant des sous-classes de produit.

enduit(P), $has_diagnostic_characteristic(P, D) \Rightarrow has_diagnostic(D, positive)$

2. Ajout d'une contrainte temporelle (année permettant d'obtenir la meilleure confiance pour $hc > minHC$).

enduit(P), *contain*(L, P), *has_location*(S, L), *has_structure*(B, S), *has_year*(B, Y), *lessThanOrEqualTo*(Y, "1970"), ...

Etapes de CRA-Miner

3. Spécialisation par localisation et/ou par structure.

enduit(P), mur(L), contain(L, P), has_location(S, L), has_structure(B, S), has_year(B, Y), lessThanOrEqual(Y, "1970"), ...

4. Enrichissement par la région.

5. Spécialisation en ajoutant d'autres composants co-localisés (les frères).

enduit(P), mur(L), peinture(P2), contain(L, P2), contain(L, P), has_location(S, L), has_structure(B, S), has_year(B, Y), lessThanOrEqual(Y, "1970"), ...

Expérimentations

Jeux de données
Analyse des résultats

Jeux de données



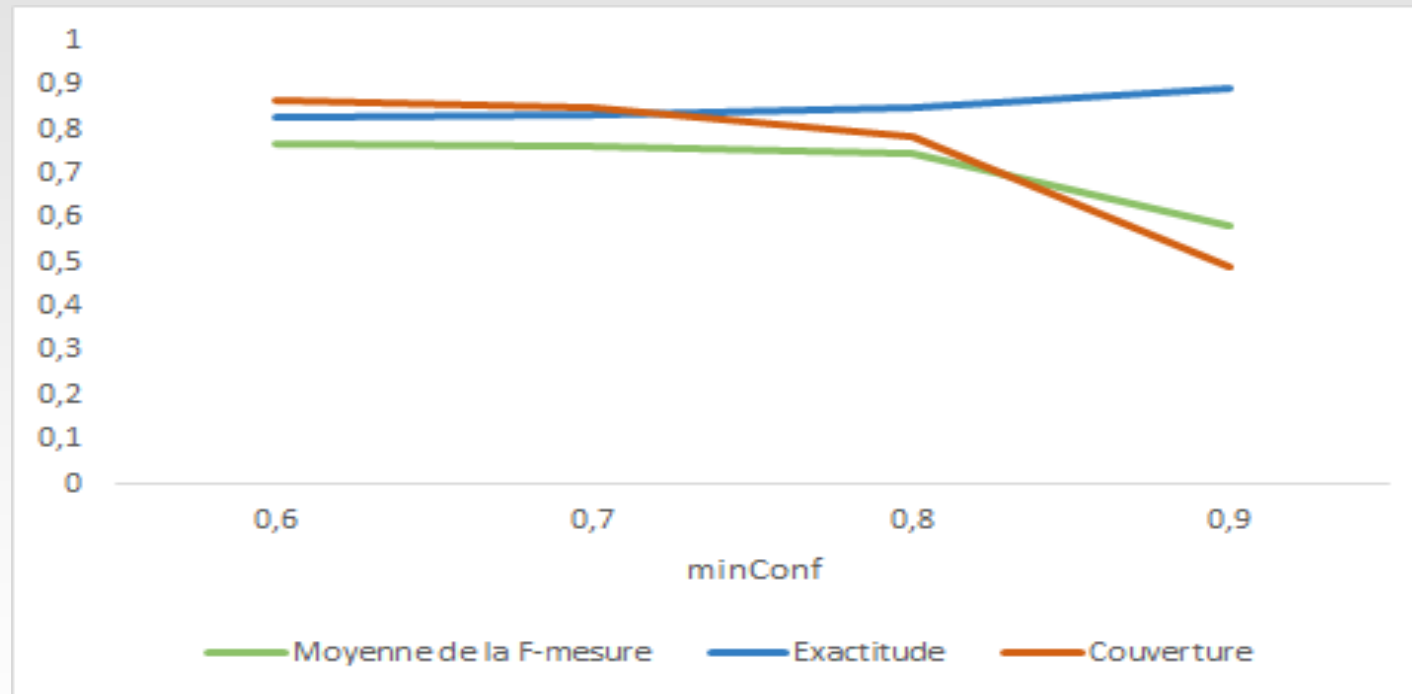
Diagnostic
amiante

- 51970 triplets.
- 2998 instances de produit, 341 localisations, 214 structures et 94 bâtiments.
- Année de construction $\in [1948,1997]$.
- 1525 **produits** contenant de l'amiante et 1473 **produits** sans amiante.

Paramètre et méthode de test

- CO={product, location, structure, contain, has_location, has_region, has_year, has_structure, has_diagnostic_characteristic }
- maxSibS = 0, maxSibL = 3, maxSibP = 3
- minHC = 0,001
- Validation croisée sur 3 tiers.
- CRA-Miner est parallélisé.

Validation de minConf = 0,6



Résultats de CRA-Miner selon différents seuils de minConf

Résultats CRA-Miner

- Approche pessimiste pour les doubles décisions.
- 29 règles comportent au moins un produit frère (au maximum 2).
- 16 règles comportent une localisation frère (au maximum 3).
- 14 règles exploitent une contrainte temporelle.

- Exemple de règle (13 atomes)

enduitsabasedeplatreprojeteslissesoutaloches(?P), Location(?L),
 Structure(?S), has_location(?S, ?L), gainesetcoffresverticaux(?L4),
 has_location(?S, ?L4),
 revetementsdemurspoteauxcloisonsgainescoffres(?L3), has_location(?S,
 ?L3), cloisonssèchesassembleeprefabriquees(?L2), has_location(?S, ?L2),
 contain(?L, ?P), has_diagnostic_characteristic(?P, ?D
 ⇒ has_Diagnosis(?D, 'positive')

minConf=0.6, minHC=0.001	CRA-Miner
# règles	75
Double décision	82
TP	415
TN	303
FP	121
FN	28
UP	66
UN	67
Précision des positives	0.77
Rappel des positives	0.82
F-mesure des positives	0.79
Précision des négatives	0.92
Rappel des négatives	0.62
F-mesure des négatives	0.74
Moyenne de la F-mesure	0.77
Exactitude	0.83
Couverture	87%

Comparaison avec une Baseline et AMIE3

minConf=0.6, minHC=0.001	CRA-Miner	AMIE3, l=4	AMIE3, l=6	Baseline
# règles	75	45	91	24
Double décision	82	50	277	0
TP	415	381	473	146
TN	303	288	264	257
FP	121	146	226	30
FN	28	74	32	24
UP	66	54	3	338
UN	67	58	0	204
Précision des positives	0.77	0.72	0.68	0.83
Rappel des positives	0.82	0.75	0.93	0.29
F-mesure des positives	0.79	0.73	0.79	0.43
Précision des négatives	0.92	0.8	0.89	0.91
Rappel des négatives	0.62	0.59	0.54	0.52
F-mesure des négatives	0.74	0.68	0.67	0.66
Moyenne de la F-mesure	0.77	0.71	0.73	0.55
Exactitude	0.83	0.75	0.74	0.88
Couverture	87%	89%	100%	46%

Comparaison de CRA-Miner avec une baseline et AMIE3 avec l=4 et l=6 (minHC=0.001, minConf=0.6)

Conclusion et perspective

- Définition d'une approche CRA-Miner qui utilise les connaissances expert, la sémantique de l'ontologie Asbestos, les informations temporelles et les heuristiques partie-tout pour découvrir des règles pertinentes.
- Les résultats montre une bonne exactitude, avec en particulier une bonne précision pour les négatifs (stratégie pessimiste).
- Nous envisageons de comparer avec d'autres approches de type Divide-and-Conquer qui utilisent un biais de langage comme TILDE [1].
- Comparer avec une approche hybride [K-Cap 2019] qui utilise des ressources externes au CSTB qui décrivent les produits commercialisés amiantés.

[1] Hendrik Blockeel and Luc De Raedt. Top-down in-duction of first-order logical decision trees. *Artificialintelligence*, 101(1-2) :285–297, 1998.

[K-Cap2019] Thamer Mecharnia, Lydia Chibout Khelifa, NathaliePernelle, and Fayçal Hamdi. An approach toward aprediction of the presence of asbestos in buildings ba-sed on incomplete temporal descriptions of marketedproducts. In Mayank Kejriwal, Pedro A. Szekely, andRaphaël Troncy, editors,K-CAP 2019, Marina DelRey, CA, USA, November 19-21, 2019, pages 239–242. ACM, 2019.

Merci