

Un cadre paraconsistant pour l'évaluation de similarité dans les bases de connaissances

José-Luis Vilchis-Medina

ENSTA - Institut Polytechnique de Paris, Lab-STICC, Robex, Brest, 29200, France

jose.vilchis@ensta.fr

Résumé

Cet article propose un **cadre paraconsistant pour l'évaluation de la similarité** dans les bases de connaissances. Contrairement aux approches classiques, ce cadre intègre explicitement les **contradictions**, permettant une mesure de similarité plus robuste et interprétable. Une nouvelle mesure S^* est introduite, pénalisant les incohérences tout en valorisant les propriétés partagées. Des **super-catégories paraconsistantes** Ξ_K^* sont définies pour organiser hiérarchiquement les entités de connaissances. Le modèle inclut également un **extracteur de contradictions** E et un mécanisme de réparation, garantissant la cohérence des évaluations. Les résultats théoriques assurent la réflexivité, la symétrie et les bornes de S^* . Cette approche offre une solution prometteuse pour la gestion des connaissances conflictuelles, avec des perspectives dans les systèmes multi-agents.

Mots-clés

Mesure de similarité, Logique paraconsistante, Contradictions, Représentation des connaissances.

Abstract

This article proposes a **paraconsistent framework for evaluating similarity** in knowledge bases. Unlike classical approaches, this framework explicitly integrates **contradictions**, enabling a more robust and interpretable similarity measure. A new measure S^* is introduced, which penalizes inconsistencies while rewarding shared properties. **Paraconsistent super-categories** Ξ_K^* are defined to hierarchically organize knowledge entities. The model also includes a **contradiction extractor** E and a repair mechanism, ensuring consistency in the evaluations. Theoretical results guarantee reflexivity, symmetry, and boundedness of S^* . This approach offers a promising solution for managing conflicting knowledge, with perspectives in multi-agent systems.

Keywords

Similarity measure, Paraconsistent logic, Contradictions, Knowledge representation.

1 Introduction

Dans un contexte marqué par une prolifération croissante d'informations, l'interaction entre les systèmes formels et

le raisonnement qualitatif s'avère essentielle [12]. La logique fournit une structure rigoureuse pour l'analyse des arguments et la résolution de problèmes complexes, tandis que le raisonnement qualitatif permet de saisir les nuances et le contexte souvent négligés par les approches purement quantitatives. Cette synergie est particulièrement cruciale dans les processus décisionnels, où les informations incomplètes ou contradictoires sont fréquentes [4].

Les bases de connaissances (BC) jouent un rôle central dans la gestion et la consultation d'informations structurées. Cependant, leur utilité est souvent entravée par des défis inhérents, tels que les contradictions, les données incomplètes et l'évolution dynamique des connaissances [17]. Les méthodes traditionnelles d'évaluation de la similarité dans les BC reposent généralement sur des approches numériques ou probabilistes, qui peinent à gérer efficacement ces problèmes [20]. En particulier, ces méthodes supposent souvent l'absence de contradiction, ce qui n'est pas réaliste dans des environnements complexes comme les systèmes juridiques ou médicaux [27]. De plus, elles présentent une rigidité hiérarchique qui limite leur adaptabilité face aux évolutions des structures de connaissance.

Pour pallier ces limitations, cet article propose un cadre fondé sur des principes de logiques paraconsistantes, un vaste ensemble de formalismes permettant de raisonner en présence de contradictions sans tomber dans la trivialité [1]. Contrairement aux logiques classiques, où une seule contradiction entraîne l'effondrement du système (principe d'explosion), les logiques paraconsistantes isolent les incohérences et permettent de continuer à raisonner sur les parties non contradictoires des connaissances [8, 21]. Bien qu'il existe de nombreuses variantes de logiques paraconsistantes, notre travail se concentre sur leur application originale au problème de la mesure de similarité dans les bases de connaissances.

Fondements de la logique paraconsistante

Parmi les différentes logiques paraconsistantes, nous adoptons un cadre symbolique basé sur des opérateurs non classiques pour modéliser les entités de connaissance. Cette approche permet de représenter explicitement les contradictions tout en conservant la cohérence globale du système. Le concept de similarité occupe une place centrale dans de nombreux domaines, notamment la médecine, le droit et l'intelligence artificielle [27]. Cependant, les mé-

thodes existantes rencontrent souvent des difficultés pour gérer les contradictions et les données incomplètes, conduisant à des évaluations erronées ou non pertinentes [22]. Notre cadre introduit la notion d'extracteur de contradictions et d'entités réparables, permettant des évaluations significatives même en présence d'incohérences. En résolvant les contradictions tout en préservant les structures hiérarchiques, cette approche garantit à la fois la cohérence et l'interprétabilité.

Cet article présente un cadre original pour surmonter les limites des mesures de similarité classiques. L'espace de propriétés de similarité Ξ_P est étendu avec des propriétés contradictoires (notées Ξ_P^*), permettant ainsi la formalisation de la similarité en présence de contradictions. La mesure de similarité paraconsistante proposée S^* quantifie la similarité en tenant compte explicitement des propriétés partagées et contradictoires, assurant une robustesse face aux données incomplètes ou conflictuelles. De plus, les super-catégories paraconsistantes Ξ_K^* fournissent une structure hiérarchique flexible pour organiser les entités de connaissances en fonction de leurs scores de similarité, améliorant ainsi l'interprétabilité et l'adaptabilité [28].

Contributions principales

Les contributions principales de ce travail incluent :

- Une définition formelle de Ξ_P^* et S^* , généralisant les mesures de similarité classiques pour intégrer les contradictions. Cette extension permet de capturer la nature nuancée des connaissances réelles, où des propriétés conflictuelles coexistent souvent.
- Des garanties théoriques, y compris la réflexivité, la symétrie de S^* , assurant des propriétés fondamentales pour les applications pratiques. Ces propriétés garantissent que la mesure de similarité reste cohérente et interprétable, même en présence de contradictions.
- Un cadre pour l'organisation hiérarchique des connaissances via Ξ_K^* , validé par des exemples issus des domaines médicaux et juridiques. Ce cadre permet de regrouper dynamiquement les entités de connaissances en fonction de leurs scores de similarité, reflétant ainsi les hiérarchies naturelles des domaines d'application.
- Un extracteur de contradictions \mathcal{E} et un mécanisme de réparation, qui permettent de gérer les incohérences tout en préservant la cohérence des bases de connaissances. Ces outils sont essentiels pour garantir la robustesse du cadre dans des contextes dynamiques et complexes.

Cet article est structuré comme suit : la Section 2 passe en revue les travaux connexes et met en évidence les limitations des approches existantes. La Section 3 introduit le cadre proposé, en détaillant les définitions formelles, les propositions et les théorèmes. Ensuite, la Section 4 analyse les contributions. Enfin, la Section 5 conclut en résumant les apports principaux et en explorant les perspectives.

2 État de l'art

Approches logiques et probabilistes

L'évaluation de la similarité dans les bases de connaissances a été largement explorée à travers diverses méthodologies. Les approches traditionnelles reposent souvent sur la logique du premier ordre (FOL) [4] ou sur des modèles probabilistes [2]. Bien que les systèmes basés sur la FOL excellent dans le raisonnement formel, ils manquent de mécanismes pour gérer les contradictions, ce qui peut entraîner une trivialisaiton dans des contextes incohérents [23]. Les méthodes probabilistes, telles que les réseaux bayésiens [7], modélisent efficacement l'incertitude mais nécessitent des hypothèses distributionnelles fortes et peuvent échouer dans des ensembles de données clairsemés ou biaisés [18]. Les progrès récents en apprentissage profond [9] et en traitement du langage naturel (NLP) [6] ont considérablement amélioré les mesures de similarité sémantique. Cependant, ces méthodes restent gourmandes en données et souvent opaques dans leurs processus de raisonnement, limitant leur interprétabilité dans des domaines critiques tels que le droit ou la médecine [15, 16].

Méthodes basées sur les graphes et les ontologies

La théorie des graphes [11] et l'alignement des ontologies [24] offrent des représentations structurées des connaissances, permettant une meilleure interprétabilité et la modélisation de relations complexes [3, 25]. Cependant, ces approches sont confrontées à des défis computationnels importants, notamment lorsqu'il s'agit de traiter de grandes bases de données. Par exemple, l'alignement des entités dans les graphes de connaissances [24] nécessite souvent des systèmes à haute mémoire et un réglage manuel approfondi. Les approches basées sur les ontologies [14] améliorent l'interprétabilité mais peinent à suivre les mises à jour dynamiques et l'évolution des connaissances, limitant ainsi leur applicabilité dans des domaines en évolution rapide [10].

Limitations des cadres existants

Les méthodes actuelles pour évaluer la similarité dans les bases de connaissances présentent plusieurs limitations critiques :

- **Ignorance des contradictions** : Les mesures de similarité classiques [22] considèrent les contradictions comme irréconciliables, conduisant à une sur-estimation de la similarité. Cette approche ne capture pas la complexité des connaissances réelles, où des propriétés conflictuelles coexistent souvent.
- **Manque d'adaptabilité hiérarchique** : Les systèmes existants imposent souvent des classifications rigides, ne reflétant pas les hiérarchies nuancées inhérentes aux domaines de connaissances. Cette rigidité limite leur capacité à s'adapter à des structures évolutives ou spécifiques à un domaine.
- **Problèmes de transparence** : Les modèles boîte noire, en particulier ceux basés sur l'apprentissage

profond [9], entravent l'interprétabilité. Ce manque de transparence est particulièrement problématique dans des domaines comme le droit ou la médecine, où la justification des décisions est essentielle.

Enfin, les métriques de similarité classiques, bien qu'efficaces dans certains contextes, souffrent de limites importantes lorsqu'il s'agit de gérer des contradictions ou des incertitudes. Des exemples notables incluent :

- La distance de Levenshtein [30], qui ne prend pas en compte les incohérences sémantiques.
- La distance de Hamming [19], qui suppose des chaînes de longueurs identiques.
- La distance de Jaro-Winkler [29], qui reste sensible aux erreurs de transcription sans gérer les conflits logiques.

La logique paraconsistante comme solution

La logique paraconsistante [5] offre une alternative prometteuse en permettant un raisonnement avec des contradictions sans s'effondrer dans la trivialité [23]. Contrairement aux logiques traditionnelles, le raisonnement paraconsistant permet des comparaisons significatives même lorsque les entités de connaissances contiennent des propriétés conflictuelles [24]. Des travaux récents ont exploré l'intégration de la logique paraconsistante avec les systèmes de représentation des connaissances, démontrant son potentiel pour améliorer la robustesse et la transparence [13, 14].

Le cadre proposé s'appuie sur cette base pour répondre aux limitations des approches existantes :

- **Intégration des contradictions** : Un extracteur de contradictions et un mécanisme de réparation sont introduits, intégrant les contradictions dans Ξ_P^* pour éviter la perte d'information et assurer la cohérence.
- **Pénalisation des contradictions** : Le cadre définit S^* pour pénaliser les contradictions, assurant une robustesse face au bruit et améliorant la précision des évaluations de similarité.
- **Regroupement hiérarchique dynamique** : En exploitant Ξ_K^* , le cadre regroupe dynamiquement les entités de connaissances via des seuils de similarité θ , reflétant les hiérarchies de domaine et permettant une adaptabilité.

Malgré ces avancées, des défis subsistent, notamment dans la gestion de l'évolution dynamique des connaissances et la nature subjective de la sélection des propriétés [26]. Le cadre proposé aborde ces problèmes en introduisant des espaces de similarité adaptatifs et des mécanismes pour une sélection cohérente des propriétés, assurant à la fois la robustesse et l'adaptabilité dans les évaluations de similarité [28].

3 Cadre proposé

Le cadre proposé repose sur une modélisation rigoureuse des entités de connaissances au sein d'un langage de logique du premier ordre, permettant l'intégration explicite des contradictions dans l'évaluation de similarité. Ce cadre inclut une mesure de similarité paraconsistante S^* , une hiérarchie organisée via des super-catégories paraconsistantes

Ξ_K^* , ainsi qu'un extracteur de contradictions E et un mécanisme de réparation pour gérer les incohérences. Les propriétés fondamentales garantissent la robustesse et l'interprétabilité du modèle.

3.1 Définitions formelles des entités de connaissance

Soit $\mathcal{K} = \{K_1, K_2, \dots, K_n\}$ l'ensemble fini des entités de connaissances traitées dans la base de connaissances. Chaque entité K_i est définie comme un ensemble fini de littéraux clos dans un langage de logique du premier ordre \mathcal{L} , soit :

$$K_i = \{\ell_1, \dots, \ell_m\}, \quad \text{où } \ell_j \in \text{Lit}(\mathcal{L}), \quad (1)$$

où $\text{Lit}(\mathcal{L})$ désigne l'ensemble des littéraux atomiques et négatifs clos de \mathcal{L} . Cette définition fournit une base sémantique claire pour évaluer les relations entre entités, tout en autorisant la coexistence de propriétés contradictoires. Chaque littéral ℓ_j peut représenter une propriété spécifique ou son négation, ce qui permet de modéliser des connaissances complexes et ambivalentes.

3.2 Espace de propriétés paraconsistantes

L'espace de propriétés de similarité classique Ξ_P est étendu pour intégrer explicitement les propriétés contradictoires, conduisant à la définition suivante :

Définition 1 (Espace paraconsistant des propriétés). On définit l'espace paraconsistant des propriétés comme :

$$\Xi_P^* = \Xi_P \cup P_{\text{contradictory}}, \quad (2)$$

où :

- Ξ_P : ensemble des propriétés partagées par deux entités K_1 et K_2 ,
- $P_{\text{contradictory}}$: ensemble des couples de propriétés $(p, \neg p)$ entre K_1 et K_2 .

Cette définition permet de formaliser les interactions entre entités de connaissances en présence de conflits, sans invalider l'ensemble du raisonnement. Elle introduit explicitement les contradictions, ce qui est crucial pour une évaluation de similarité robuste en contextes incertains ou conflictuels.

3.3 Mesure de similarité paraconsistante S^*

La mesure S^* est reformulée pour refléter à la fois la similarité positive et la divergence induite par les contradictions :

Définition 2 (Mesure de similarité paraconsistante). Soient K_1 et K_2 deux entités de connaissances. On définit la mesure de similarité paraconsistante S^* comme :

$$S^*(K_1, K_2) = S^+(K_1, K_2) - D^\pm(K_1, K_2), \quad (3)$$

où :

$$S^+(K_1, K_2) = \frac{|P_{\text{shared}}|}{|P_{\text{total}}|}, \quad D^\pm(K_1, K_2) = \frac{|P_{\text{contradictory}}|}{|P_{\text{total}}|}. \quad (4)$$

Ainsi, $S^* \in [-1, 1]$, avec :

- $S^* > 0$: mesures nettement similaires,
- $S^* = 0$: équilibre entre similitude et divergence,
- $S^* < 0$: divergence dominante.

Cette formulation permet de pénaliser les contradictions tout en valorisant les propriétés partagées. Par exemple, si K_1 et K_2 partagent une propriété p mais contiennent également une contradiction $(p, \neg p)$, la mesure S^* sera négative, reflétant une divergence nette. Cela garantit une interprétation claire et intuitive de la similarité en présence de conflits.

3.4 Propriétés fondamentales de S^*

Les propriétés essentielles garantissant la cohérence du cadre sont les suivantes :

Proposition 1 (Réflexivité). Pour toute entité de connaissance K , on a :

$$S^*(K, K) = 1. \quad (5)$$

Démonstration. Par définition, $P_{\text{shared}} = P_{\text{total}}$ et $P_{\text{contradictory}} = \emptyset$. Donc :

$$S^*(K, K) = \frac{|P_{\text{total}}|}{|P_{\text{total}}|} - \frac{0}{|P_{\text{total}}|} = 1 - 0 = 1. \quad (6)$$

□

Proposition 2 (Symétrie). Pour tous $K_1, K_2 \in \mathcal{K}$, on a :

$$S^*(K_1, K_2) = S^*(K_2, K_1). \quad (7)$$

Proposition 3 (Bornes extrêmes). Pour toutes entités K_1, K_2 , la mesure S^* satisfait :

$$-1 \leq S^*(K_1, K_2) \leq 1. \quad (8)$$

Ces propriétés garantissent que S^* est une mesure cohérente et interprétable, même en présence de contradictions. La réflexivité et la symétrie sont des propriétés standard pour les mesures de similarité, tandis que les bornes garantissent que S^* reste dans un intervalle raisonnable.

3.5 Super-catégories paraconsistantes Ξ_K^*

Les entités de connaissances peuvent être organisées hiérarchiquement selon leur niveau de similarité via le seuil $\theta \in [-1, 1]$.

Définition 3 (Super-catégories paraconsistantes). Étant donné un seuil θ , les super-catégories paraconsistantes forment un ensemble défini par :

$$\Xi_K^* = \bigcup_{i=1}^n \{K_i \mid S^*(K_i, K_j) > \theta, \forall j \neq i\}. \quad (9)$$

Théorème 1 (Disjonction des super-catégories). Si $K_1 \in \Xi_K^*$ et $K_2 \in \Xi_K^*$ appartiennent à différentes super-catégories, alors :

$$S^*(K_1, K_2) \leq \theta. \quad (10)$$

Les super-catégories Ξ_K^* permettent de regrouper dynamiquement les entités de connaissances en fonction de leur score de similarité, tout en respectant une hiérarchie naturelle. Cela améliore l'interprétabilité et l'adaptabilité du cadre, en particulier dans des domaines de connaissances dynamiques.

3.6 Extracteur de contradictions et mécanisme de réparation

L'un des défis principaux dans la gestion des bases de connaissances est la présence de contradictions, qui peuvent compromettre la cohérence et la fiabilité des analyses. Pour surmonter ce problème, le cadre proposé inclut un mécanisme de gestion des contradictions, composé d'un extracteur de contradictions E et d'un mécanisme de réparation. Ces outils permettent de détecter, résoudre et gérer les incohérences tout en préservant la structure hiérarchique des connaissances.

3.6.1 Définition de l'extracteur de contradictions

L'extracteur de contradictions E est un opérateur logique qui identifie les propriétés contradictoires présentes dans une entité de connaissance K . Formellement, on définit :

Définition 4 (Extracteur de contradictions). Un extracteur de contradictions est un opérateur logique E tel que :

$$E(K) = \{p_i \in K \mid \exists q_j \in K \text{ tels que } (p_i \wedge \neg q_j) \vee (\neg p_i \wedge q_j)\}. \quad (11)$$

L'extracteur de contradictions E identifie les propriétés p_i dans K qui sont en contradiction avec d'autres propriétés q_j de la même entité. Cela permet de localiser les sources de conflits dans les bases de connaissances. Par exemple, si une entité K contient à la fois p et $\neg p$, alors p sera identifié comme contradictoire par E .

3.6.2 Entités réparables

Une entité de connaissance peut être réparée si certaines contradictions peuvent être résolues sans affecter les autres propriétés. Cette notion est formalisée comme suit :

Définition 5 (Entité réparable). Une entité K est réparable si $\exists R \subseteq E(K)$ tel que :

$$\neg \text{Contradictoire}(K \setminus R). \quad (12)$$

Une entité K est réparable si elle contient un sous-ensemble R de contradictions qui, une fois supprimé ou résolu, permet de rendre K cohérente. Cela garantit que les bases de connaissances peuvent être maintenues cohérentes, même en présence de conflits. Par exemple, si une entité K contient une contradiction $(p, \neg p)$ mais que la suppression de $\neg p$ suffit à rendre K cohérente, alors K est réparable.

3.6.3 Mécanisme de réparation minimale

Le mécanisme de réparation vise à identifier le sous-ensemble minimal de contradictions à résoudre pour rendre une entité cohérente. Cela est formalisé par le corollaire suivant :

Corollaire 1.1 (Réparation minimale). La réparation minimale R_{\min} est définie comme :

$$R_{\min} = \arg \min_{R \subseteq E(K)} (\|R\| \mid \neg \text{Contradictoire}(K \setminus R)). \quad (13)$$

La réparation minimale R_{\min} est le sous-ensemble de contradictions le plus petit qui, une fois résolu, permet de rendre K cohérente. Cela garantit que le nombre de modifications apportées à K est minimisé, ce qui est essentiel pour préserver l'intégrité des connaissances. Par exemple, si une entité K contient plusieurs contradictions, la réparation minimale R_{\min} consiste à résoudre uniquement celles qui sont nécessaires pour rendre K cohérente.

3.6.4 Impact sur la mesure de similarité paraconsistante

L'introduction de l'extracteur de contradictions et du mécanisme de réparation permet de définir une version améliorée de la mesure de similarité paraconsistante, qui tient compte des réparations effectuées :

Définition 6 (Similarité paraconsistante avec réparations). Soient K_1 et K_2 deux entités de connaissances. La similarité paraconsistante avec réparations $\Xi_{RP}(K_1, K_2)$ est définie comme :

$$\Xi_{RP}(K_1, K_2) = \Xi_P(K'_1, K'_2), \quad (14)$$

où K'_1 et K'_2 sont les versions réparées de K_1 et K_2 , respectivement, obtenues en résolvant les contradictions à l'aide de E .

La similarité paraconsistante avec réparations Ξ_{RP} permet de comparer les entités de connaissances après que les contradictions ont été résolues. Cela garantit une évaluation plus précise et cohérente de la similarité, même en présence de conflits. Par exemple, si K_1 et K_2 contiennent des contradictions, la similarité $\Xi_{RP}(K_1, K_2)$ est calculée sur les versions réparées K'_1 et K'_2 , ce qui réduit l'impact des contradictions sur la mesure de similarité.

3.6.5 Propriétés théoriques

Le cadre proposé garantit plusieurs propriétés théoriques essentielles pour assurer la cohérence et la robustesse des évaluations de similarité, même en présence de contradictions.

Proposition 4 (Existence de l'extracteur de contradictions). Pour toute entité de connaissance K , il existe un extracteur de contradictions $E(K)$.

Démonstration. D'après la définition 4, $E(K)$ est construit en identifiant des paires de propriétés p_i et q_j telles que $p_i \wedge \neg q_j$ ou $\neg p_i \wedge q_j$. Puisque la logique du premier ordre (FOL) garantit que les propriétés peuvent être exprimées logiquement, $E(K)$ est bien défini. \square

Proposition 5 (Réparabilité d'une entité). Une entité de connaissance K est réparable si et seulement si $E(K) \neq K$.

Démonstration. Si $E(K) = K$, alors chaque propriété de K est impliquée dans une contradiction, rendant K irréparable. Inversement, si $E(K) \neq K$, il existe au moins une propriété non impliquée dans une contradiction, permettant l'existence d'un sous-ensemble cohérent $K \setminus R$. \square

Théorème 2 (Préservation des catégories de similarité). La similarité paraconsistante avec réparations $\Xi_{RP}(K_1, K_2)$ préserve les catégories de similarité originales $K' \approx, K' \neq$ après réparation de K_1 et K_2 .

Démonstration. D'après la définition 6, $\Xi_{RP}(K_1, K_2)$ est calculée sur les versions réparées K'_1 et K'_2 . Puisque les réparations résolvent les contradictions sans altérer les propriétés partagées, les catégories de similarité restent cohérentes avec l'original $\Xi_P(K_1, K_2)$. \square

Théorème 3 (Cohérence après réparation). Si S est une super-catégorie dans Ξ_K , alors la réparation de S assure la cohérence dans toutes les sous-catégories.

Démonstration. D'après la définition 6, la réparation de S implique la résolution des contradictions dans ses propriétés constitutives. Puisque les sous-catégories héritent des propriétés de S , la réparation de S assure la cohérence à tous les niveaux de la hiérarchie. \square

Corollaire 3.1 (Réparation minimale). La réparation minimale R_{\min} satisfait :

$$R_{\min} = \arg \min_{R \subseteq E(K)} (\|R\| \mid \neg \text{Contradictoire}(K \setminus R)). \quad (15)$$

Démonstration. D'après la proposition 5, R_{\min} est le plus petit sous-ensemble de $E(K)$ qui résout toutes les contradictions, assurant la réparabilité. \square

Corollaire 3.2 (Préservation de la structure hiérarchique). La réparation d'une super-catégorie S n'affecte pas la structure hiérarchique de Ξ_K .

Démonstration. D'après le théorème 3, la réparation de S assure la cohérence dans les sous-catégories, préservant l'organisation hiérarchique de Ξ_K . \square

3.7 Exemples

Nous illustrons ici le fonctionnement du cadre avec deux exemples concrets :

Exemple 1 (Contradictions simples). Soient :

$$K_1 = \{p_1, p_2, \neg p_3\}, \quad K_2 = \{p_2, p_3, \neg p_1\}. \quad (16)$$

Alors :

$$\begin{aligned} P_{\text{shared}} &= \{p_2\}, \\ P_{\text{contradictory}} &= \{p_1, p_3\}, \\ P_{\text{total}} &= \{p_1, p_2, p_3, \neg p_1, \neg p_3\}. \end{aligned}$$

Donc :

$$S^*(K_1, K_2) = \frac{1}{5} - \frac{2}{5} = -\frac{1}{5}.$$

Cela montre que malgré une propriété partagée, les contradictions entraînent une valeur de similarité négative.

Dans cet exemple, les entités K_1 et K_2 partagent une seule propriété p_2 , mais présentent deux contradictions ($p_1, \neg p_1$) et ($p_3, \neg p_3$). La mesure S^* prend en compte ces contradictions, ce qui entraîne une valeur négative. Cela illustre comment les contradictions peuvent réduire significativement la similarité entre deux entités, reflétant ainsi la complexité des connaissances réelles.

Exemple 2 (Regroupement hiérarchique). Considérons un ensemble de cinq entités de connaissances $\mathcal{K} = \{K_1, K_2, K_3, K_4, K_5\}$, chacune représentant une base de connaissances concernant des diagnostics médicaux. Les entités sont définies comme suit :

$$\begin{aligned} K_1 &= \{\text{fièvre, toux, } \neg\text{maux de tête}\}, \\ K_2 &= \{\text{fièvre, } \neg\text{toux, maux de tête}\}, \\ K_3 &= \{\text{fièvre, toux, fatigue}\}, \\ K_4 &= \{\text{essoufflement, vomissements, } \neg\text{fièvre}\}, \\ K_5 &= \{\text{essoufflement, } \neg\text{vomissements, douleur abdominale}\}. \end{aligned}$$

Étape 1 : Calcul des mesures de similarité paraconsistante S^*

Nous appliquons la mesure de similarité paraconsistante S^* entre chaque paire d'entités, définie par l'Équation 3. Voici les résultats calculés :

$$\begin{aligned} S^*(K_1, K_2) &= \frac{1-2}{5} = -0.2, \\ S^*(K_1, K_3) &= \frac{2-0}{4} = 0.5, \\ S^*(K_1, K_4) &= \frac{0-1}{6} = -0.17, \\ S^*(K_1, K_5) &= \frac{0-0}{5} = 0, \\ S^*(K_2, K_3) &= \frac{1-1}{5} = 0, \\ S^*(K_2, K_4) &= \frac{0-1}{6} = -0.17, \\ S^*(K_2, K_5) &= \frac{0-0}{5} = 0, \\ S^*(K_3, K_4) &= \frac{0-1}{6} = -0.17, \\ S^*(K_3, K_5) &= \frac{0-0}{5} = 0, \\ S^*(K_4, K_5) &= \frac{1-1}{5} = 0. \end{aligned}$$

Étape 2 : Définition du seuil de similarité θ

Fixons un seuil $\theta = 0.4$. Selon la définition des super-catégories paraconsistantes (cf. section 3.5), deux entités appartiennent à la même super-catégorie si leur similarité dépasse ce seuil.

Étape 3 : Formation des super-catégories paraconsistantes Ξ_K^*

À partir des valeurs ci-dessus, nous identifions les paires satisfaisant $S^*(K_i, K_j) > \theta$:

- $S^*(K_1, K_3) = 0.5 > 0.4$, donc K_1 et K_3 appartiennent à la même super-catégorie.
- Toutes les autres paires ont $S^* \leq 0.4$, donc restent isolées.

Ainsi, les super-catégories paraconsistantes sont :

$$\Xi_K^* = \{\{K_1, K_3\}, \{K_2\}, \{K_4\}, \{K_5\}\}. \quad (17)$$

Étape 4 : Gestion des contradictions via l'extracteur E

Examinons K_1 et K_2 , dont la similarité est négative ($S^* = -0.2$). Appliquons l'extracteur de contradictions E sur ces deux entités :

$$E(K_1) = \emptyset, \quad E(K_2) = \{\text{toux, } \neg\text{toux}\}. \quad (18)$$

Supprimons $\neg\text{toux}$ de K_2 pour obtenir une version réparée $K_2' = \{\text{fièvre, maux de tête}\}$, puis recalculons la similarité :

$$S^*(K_1, K_2') = \frac{1-0}{4} = 0.25 < \theta. \quad (19)$$

Même après réparation, K_1 et K_2' ne dépassent pas le seuil de regroupement.

Étape 5 : Hiérarchie finale et interprétation

La structure hiérarchique finale est donc :

$$\Xi_K^* = \{\{K_1, K_3\}, \{K_2\}, \{K_4\}, \{K_5\}\}. \quad (20)$$

L'interprétation clinique est claire :

- $\{K_1, K_3\}$ représente des patients présentant des symptômes similaires liés aux voies respiratoires supérieures.
- K_2 , malgré sa proximité sémantique, reste isolé en raison des contradictions persistantes.
- $\{K_4, K_5\}$, bien que partageant une propriété commune (essoufflement), sont séparés car leurs différences symptomatiques dominent.

Cet exemple illustre comment le cadre paraconsistant permet de structurer des bases de connaissances complexes en tenant compte non seulement des similitudes, mais aussi des contradictions. La mesure S^* , couplée à l'extracteur de contradictions E , garantit une évaluation robuste et interprétable, même en présence de conflits logiques.

4 Discussion

Le cadre proposé pour l'évaluation de la similarité dans les bases de connaissances représente une avancée significative par rapport aux méthodes classiques. En intégrant explicitement les contradictions via la logique paraconsistante, il permet une évaluation plus précise, interprétable et robuste de la similarité entre entités de connaissances. Cette section analyse les contributions principales du cadre, compare sa mesure S^* aux approches traditionnelles, souligne ses avantages théoriques et pratiques, et identifie des perspectives d'amélioration future.

4.1 Contributions principales

Le cadre introduit une mesure de similarité paraconsistante S^* , définie par l'Équation 3 :

$$S^*(K_1, K_2) = S^+(K_1, K_2) - D^\pm(K_1, K_2),$$

Cette mesure étend les mesures classiques en intégrant les contradictions de manière explicite, ce qui permet de pénaliser les divergences tout en valorisant les similitudes.

En outre, le cadre propose des **super-catégories paraconsistantes** Ξ_K^* , définies par l'Équation 9 :

$$\Xi_K^* = \bigcup_{i=1}^n \{K_i \mid S^*(K_i, K_j) > \theta, \forall j \neq i\}.$$

Ces super-catégories organisent hiérarchiquement les entités de connaissances selon leur niveau de similarité, améliorant ainsi l'interprétabilité et l'adaptabilité du modèle.

Un mécanisme central du cadre est également l'extracteur de contradictions E , défini par l'Équation 11 :

$$E(K) = \{p_i \in K \mid \exists q_j \in K \text{ tels que } (p_i \wedge \neg q_j) \vee (\neg p_i \wedge q_j)\}.$$

L'extracteur E permet d'identifier les incohérences locales, facilitant ainsi la gestion des conflits sans compromettre la cohérence globale du système.

4.2 Avantages par rapport aux approches existantes

Le cadre proposé surmonte plusieurs limitations des méthodes classiques :

- **Prise en compte des contradictions** : Contrairement aux méthodes classiques qui ignorent ou neutralisent les contradictions, le cadre paraconsistant les intègre dans l'évaluation, ce qui permet une mesure de similarité plus fidèle à la réalité des données conflictuelles.
- **Flexibilité hiérarchique** : Les super-catégories paraconsistantes offrent une organisation dynamique basée sur un seuil variable θ , contrairement aux classifications rigides des approches traditionnelles.
- **Transparence accrue** : Le cadre repose sur des principes logiques explicites, ce qui le distingue des modèles boîte noire comme ceux basés sur l'apprentissage profond. Cela est crucial dans des domaines critiques comme la médecine ou le droit.

Comparaison avec des mesures classiques

Pour illustrer la valeur ajoutée de la mesure de similarité paraconsistante S^* par rapport aux approches traditionnelles, nous présentons ici une comparaison explicite avec la mesure de Jaccard, souvent utilisée pour quantifier la similarité entre ensembles.

Définition 7 (Similarité de Jaccard). Étant donnés deux ensembles finis A et B , la similarité de Jaccard est donnée par :

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Bien que cette mesure soit robuste et intuitive dans les cas non contradictoires, elle ne prend pas en compte les propriétés conflictuelles, ce qui peut mener à des évaluations erronées dans des contextes où les contradictions sont présentes.

Exemple : S^* vs. Jaccard

Considérons deux entités de connaissances K_1 et K_2 , définies comme suit :

$$K_1 = \{p_1, p_2, \neg p_3\}, \quad K_2 = \{p_2, p_3, \neg p_1\}.$$

Les ensembles de propriétés associés sont :

- $P_{\text{shared}} = \{p_2\}$,
- $P_{\text{contradictory}} = \{p_1, p_3\}$,
- $P_{\text{total}} = \{p_1, p_2, p_3, \neg p_1, \neg p_3\}$.

Calculons à la fois la mesure paraconsistante S^* et la mesure de Jaccard :

- Mesure paraconsistante S^* :

$$\begin{aligned} S^*(K_1, K_2) &= \frac{|P_{\text{shared}}| - |P_{\text{contradictory}}|}{|P_{\text{total}}|} \\ &= \frac{1 - 2}{5} = -\frac{1}{5} = -0.2. \end{aligned}$$

- Mesure de Jaccard : En considérant uniquement les littéraux positifs (approche standard), on a :

$$A = \{p_1, p_2\}, \quad B = \{p_2, p_3\},$$

donc :

$$J(A, B) = \frac{|\{p_2\}|}{|\{p_1, p_2, p_3\}|} = \frac{1}{3} \approx 0.33.$$

Dans cet exemple, la mesure de Jaccard attribue une valeur positive à la similarité (~ 0.33), suggérant une certaine proximité entre les deux entités. Cependant, cette évaluation ignore complètement les contradictions présentes entre p_1 et $\neg p_1$, ainsi que p_3 et $\neg p_3$, ce qui peut conduire à une surestimation de la similarité dans des contextes conflictuels. En revanche, la mesure S^* intègre explicitement ces contradictions, réduisant la similarité globale et produisant une valeur négative (-0.2). Ce résultat reflète plus fidèlement la réalité : malgré une propriété partagée, les incohérences dominent, rendant la similarité nettement problématique.

L'exemple présenté précédemment illustre clairement l'avantage théorique et pratique de la mesure S^* par rapport aux mesures classiques comme celle de Jaccard. En intégrant les contradictions dans l'évaluation, S^* offre une vision nuancée, interprétable et robuste de la similarité entre entités de connaissances, particulièrement utile dans des domaines tels que le droit, la médecine ou les systèmes multi-agents où la gestion des conflits est cruciale.

Cette capacité à distinguer les similitudes réelles des faux positifs dus à des incohérences constitue une contribution essentielle du cadre paraconsistant proposé.

4.3 Perspectives futures

Malgré les avantages de ce cadre, il présente des défis qu'il conviendrait d'explorer dans le cadre de recherches futures :

- **Complexité computationnelle** : L'identification des contradictions et le calcul de S^* peuvent devenir coûteux dans de grandes bases de connaissances. Des algorithmes optimisés ou des techniques d'approximation pourraient être développés.
- **Adaptation aux bases de connaissances évolutives** : Une extension du cadre pourrait inclure des mécanismes pour gérer l'évolution dynamique des connaissances, par exemple en intégrant des mises à jour incrémentielles ou des techniques d'apprentissage en ligne.
- **Automatisation de la sélection des propriétés** : Actuellement, la sélection des propriétés est manuelle. Des méthodes automatisées basées sur l'apprentissage supervisé ou non supervisé pourraient améliorer l'objectivité et la pertinence des comparaisons.
- **Validation empirique** : Bien que des exemples illustratifs aient été fournis, une évaluation sur des ensembles de données réels restent à réaliser pour confirmer l'efficacité du cadre dans des applications pratiques.

Applications dans les systèmes multi-agents

Le cadre paraconsistant proposé peut être appliqué dans les systèmes multi-agents (SMA), notamment pour la résolution de conflits entre agents ayant des représentations de connaissances hétérogènes ou conflictuelles. Par exemple, les agents pourraient utiliser S^* pour détecter et résoudre les incohérences dans leurs bases de connaissances partagées, facilitant ainsi une coopération plus robuste et cohérente.

De plus, les super-catégories paraconsistantes Ξ_K^* pourraient servir à organiser dynamiquement les relations entre agents, en fonction de seuils de similarité θ ajustables. Cette flexibilité est particulièrement précieuse dans des environnements dynamiques comme les réseaux de capteurs distribués ou les systèmes autonomes.

Intégration potentielle dans des formalismes logiques étendus

Une voie prometteuse serait l'intégration du cadre paraconsistant avec d'autres formalismes logiques, tels que :

- **Logique modale** : Pour modéliser la croyance, la possibilité ou la nécessité dans les évaluations de similarité.
- **Logique temporelle** : Pour prendre en compte l'évolution des connaissances dans le temps.
- **Ontologies et graphes de connaissances** : Pour renforcer l'interopérabilité avec les standards actuels de représentation des connaissances.

Ces extensions pourraient enrichir les capacités du cadre tout en conservant son fondement paraconsistant.

5 Conclusion

Le présent travail a proposé un **cadre paraconsistant pour l'évaluation de la similarité dans les bases de connaissances**, visant à surmonter les limitations des approches classiques. Ce cadre s'appuie sur une **logique paraconsistante**, permettant de gérer efficacement les contradictions tout en préservant la cohérence et l'interprétabilité des évaluations de similarité.

Une **mesure de similarité paraconsistante** S^* a été introduite, qui intègre explicitement les propriétés contradictoires, offrant ainsi une évaluation plus robuste et nuancée que les mesures traditionnelles. Cette mesure a été formalisée en tenant compte des propriétés partagées et contradictoires, assurant une réflexivité et une symétrie, des propriétés essentielles pour une application pratique.

Un **ensemble de super-catégories paraconsistantes** Ξ_K^* a également été défini, permettant d'organiser hiérarchiquement les entités de connaissances selon leurs scores de similarité. Cette structure a été validée par des exemples illustratifs dans des domaines tels que la médecine et le droit, montrant sa pertinence pratique.

Un **extracteur de contradictions** E a été intégré au cadre, permettant d'identifier et de résoudre les incohérences locales, tout en préservant la cohérence globale des bases de connaissances. Des garanties théoriques ont été fournies, notamment sur la réparabilité des entités et la préservation de la structure hiérarchique après réparation.

L'analyse comparative avec des méthodes classiques, telles que la mesure de Jaccard, a permis de mettre en évidence les avantages du cadre paraconsistant : une gestion explicite des contradictions, une flexibilité accrue dans l'organisation des connaissances, et une meilleure interprétabilité des résultats.

Cependant, plusieurs défis subsistaient, notamment la **complexité computationnelle** lors de l'évaluation à grande échelle, l'**évolution dynamique** des bases de connaissances, et la **sélection des propriétés** pour la comparaison. Ces limites ont été identifiées comme des axes de recherche futurs, ouvrant la voie à des améliorations potentielles, comme l'intégration d'algorithmes optimisés, de mécanismes d'adaptation dynamique, ou de techniques d'apprentissage automatique pour la sélection des propriétés.

Enfin, une **perspective prometteuse** a été soulignée concernant l'intégration de ce cadre dans les **systèmes multi-agents (SMA)**, où la gestion des conflits et la coordination entre agents sont cruciales. Le cadre proposé a démontré son potentiel pour améliorer la robustesse, la transparence et l'adaptabilité des systèmes basés sur la logique paraconsistante.

Remerciements

L'auteur tient à remercier les personnes qui ont soutenu ce travail, ainsi que les relecteurs anonymes pour leurs commentaires constructifs qui ont contribué à l'amélioration de cet article.

Références

- [1] Jean-Yves Béziau. Are paraconsistent negations negations? In *Paraconsistency*, pages 465–486. CRC Press, 2002.
- [2] Veronica Biazzo, Angelo Gilio, Thomas Lukasiewicz, and Giuseppe Sanfilippo. Probabilistic logic under coherence, model-theoretic probabilistic logic, and default reasoning in system p. *Journal of Applied Non-Classical Logics*, 12(2) :189–213, 2002.
- [3] Vincent D Blondel, Anahí Gajardo, Maureen Heymans, Pierre Senellart, and Paul Van Dooren. A measure of similarity between graph vertices : Applications to synonym extraction and web searching. *SIAM review*, 46(4) :647–666, 2004.
- [4] Ronald Brachman and Hector Levesque. *Knowledge representation and reasoning*. Elsevier, 2004.
- [5] Ronald J Brachman and Hector J Levesque. The tractability of subsumption in frame-based description languages. In *AAAI*, volume 84, pages 34–37, 1984.
- [6] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. Universal sentence encoder. *arXiv preprint arXiv :1803.11175*, 2018.
- [7] Huanhuan Cheng and Runsheng Wang. Semantic modeling of natural scenes based on contextual bayesian networks. *Pattern Recognition*, 43(12) :4042–4054, 2010.
- [8] Newton CA Da Costa. On the theory of inconsistent formal systems. *Notre dame journal of formal logic*, 15(4) :497–510, 1974.
- [9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert : Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics : human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.
- [10] AnHai Doan, Jayant Madhavan, Pedro Domingos, and Alon Halevy. Ontology matching : A machine learning approach. *Handbook on ontologies*, pages 385–403, 2004.
- [11] Xinbo Gao, Bing Xiao, Dacheng Tao, and Xuelong Li. A survey of graph edit distance. *Pattern Analysis and applications*, 13 :113–129, 2010.
- [12] R.L. Goldstone and J.Y. Son. *Similarity*. Oxford University Press, 2012.
- [13] Benjamin N Grosz, Ian Horrocks, Raphael Volz, and Stefan Decker. Description logic programs : Combining logic programs with description logic. In *Proceedings of the 12th international conference on World Wide Web*, pages 48–57, 2003.
- [14] Nicola Guarino and Christopher A Welty. An overview of ontoclean. *Handbook on ontologies*, pages 201–220, 2009.
- [15] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57, 1999.
- [16] Jaz Kandola, Nello Cristianini, and John Shawe-taylor. Learning semantic similarity. *Advances in neural information processing systems*, 15, 2002.
- [17] Leonid Libkin. Incomplete data : what went wrong, and how to fix it. In *Proceedings of the 33rd ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 1–13, 2014.
- [18] Eliana Minicozzi. Some natural properties of strong-identification in inductive inference. *Theoretical Computer Science*, 2(3) :345–360, 1976.
- [19] Mohammad Norouzi, David J Fleet, and Russ R Salakhutdinov. Hamming distance metric learning. *Advances in neural information processing systems*, 25, 2012.
- [20] Judea Pearl. *Probabilistic reasoning in intelligent systems : networks of plausible inference*. Elsevier, 2014.
- [21] Graham Priest. *Beyond the limits of thought*. Oxford University Press, 2002.
- [22] Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*, 1995.
- [23] Mikhail Sheremet, Dmitry Tishkovsky, Frank Wolter, and Michael Zakharyashev. A logic for concepts and similarity. *Journal of Logic and Computation*, 17(3) :415–452, 2007.
- [24] Pavel Shvaiko and Jérôme Euzenat. Ontology matching : state of the art and future challenges. *IEEE Transactions on knowledge and data engineering*, 25(1) :158–176, 2011.
- [25] John F Sowa. Conceptual graphs. *Foundations of artificial intelligence*, 3 :213–237, 2008.
- [26] Peter D Turney and Patrick Pantel. From frequency to meaning : Vector space models of semantics. *Journal of artificial intelligence research*, 37 :141–188, 2010.
- [27] Amos Tversky. Features of similarity. *Psychological review*, 84(4) :327, 1977.
- [28] José-Luis Vilchis-Medina. Building intelligent databases through similarity : Interaction of logical and qualitative reasoning. In *2024 7th International Conference on Algorithms, Computing and Artificial Intelligence (ACAI)*, pages 1–5, 2024.
- [29] Yaoshu Wang, Jianbin Qin, and Wei Wang. Efficient approximate entity matching using jaro-winkler distance. In *International conference on web information systems engineering*, pages 231–239. Springer, 2017.
- [30] Li Yujian and Liu Bo. A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, 29(6) :1091–1095, 2007.