OntoFiC : une ontologie pour la détection de fraude financière et la modélisation des comportements des clients

Lylia Abrouk [1], **Hamza Chergui** [1,2], **Hamid Ahaggach** [1], **Benjamin Auger** [1], **Dominique Cheron** [2]

1. Laboratoire d'Informatique de Bourgogne - EA 7534
Université de Bourgogne
9, avenue Alain Savary, F-21078 Dijon - France
prénom.nom@u-bourgogne.fr

2. SKAIZen Group
hchergui,dcheron@skaizengroup.fr

RÉSUMÉ. *La détection de fraude est un problème complexe pour les institutions financières. Dans cet article, nous présentons notre approche pour détecter les transactions frauduleuses dans le réseau SWIFT basée sur une ontologie du domaine. Nous présentons l'ontologie OntoFiC construite pour la modélisation des transactions et des acteurs SWIFT. Cette ontologie est peuplée avec un ensemble de données réelles. Nous avons développé une approche basée sur des règles associées à des scénarios de fraude.*

ABSTRACT. *Fraud detection is a complex issue for financial institutions. In this article, we present our approach to detect fraudulent transactions in SWIFT network based on the domain ontology. We present the OntoFiC ontology constructed for the modeling of SWIFT transactions and actors. This ontology is populated with a real dataset. We developed rules-based approach with rules associated to fraud scenarios.*

MOTS-CLÉS : *Ontologie, raisonnement, détection de fraude, finance*

KEYWORDS: *Ontology, reasoning, fraud detection, finance*

1. Introduction

La détection de fraude est un problème complexe et constitue un véritable challenge pour les institutions financières. Les transactions frauduleuses ne sont pas fréquentes, mais les conséquences sur ces institutions peuvent aller du remboursement de la somme au client jusqu'à des amendes importantes dans les cas par exemple de blanchiment d'argent. Pour cela, les institutions financières doivent disposer d'outils

pour la prévention et détection de fraudes. L'entreprise SKAIZen Group développe un projet de recherche et d'innovation qui a pour but de modéliser les clients et les institutions financières en peuplant une base de connaissances à partir de différentes sources de données. Cette base permettra d'optimiser les moteurs de détection de fraudes dans les transactions financières. Le travail présenté dans cet article s'inscrit dans le cadre de notre collaboration avec SKAIZen Group dans le projet France relance (Auger *et al*., 2022) qui a pour objectif la construction d'une base de connaissances alimentée par des données transactionnelles en prenant en compte les données financières par type (client, compte, institution financière) et par relation (bénéficiaire, débiteur, compte/client, etc.). Nous proposons dans ce travail une ontologie spécialisée sur les informations financières modélisant les informations des transactions bancaires dans le réseau SWIFT et les informations des acteurs : clients et institutions financières. Ce travail est basé sur le modèle SWIFT qui est devenu une norme ISO 20022 [1] et notre base de connaissances clients KYC (Jabbari *et al*., 2020). Cette ontologie est peuplée à partir de transactions bancaires de sources hétérogènes. La création d'une ontologie spécialisée sur les informations financières permet d'une part d'interroger la base de connaissances sur les relations entre les différentes entités (personnes ou organisations) et, d'autre part, sur la détection de fraudes à partir de règles.

La suite de l'article est organisée de la manière suivante : dans la section 2, nous dressons un état de l'art des techniques sémantiques de détection de fraudes financières basées sur les ontologies. Nous proposons notre approche dans la section 3. Afin de valider notre approche, nous avons réalisé des expérimentations à partir d'un jeu de données réel. Nous concluons notre travail dans la section 4 et nous donnons quelques perspectives.

2. État de l'art : Les techniques basées sur la sémantique

Les ontologies sont utilisées depuis plusieurs années pour la représentation des connaissances dans plusieurs domaines (commerce, santé, biologie, financier). Ces dernières années, plusieurs travaux ont utilisé les ontologies dans le domaine financier pour la représentation des connaissances et la détection des fraudes financières. L'ontologie permet d'une part la représentation du domaine et la possibilité de définir des règles afin d'analyser les transactions et ainsi de détecter des fraudes ou des comportements suspicieux. (Attigeri *et al*., 2018) proposent une ontologie pour modéliser les fraudes bancaires. La méthode TF-IDF est utilisée pour trouver les phrases où l'expérience de fraude apparaît le plus. Le modèle Latent Dirichlet Allocation (LDA) est utilisé afin d'extraire les termes clés du domaine pour créer l'ontologie avec le thésaurus WordNet et définir les relations. L'ontologie est utilisée comme référence afin de détecter les nouvelles transactions frauduleuses. (El Orche *et al*., 2018) proposent une approche pour prévenir et détecter des transactions frauduleuses dans des systèmes de paiement électroniques. L'approche est basée sur une ontologie contenant les acti-

1. https://www.iso20022.org/iso-20022-message-definitions

vités, les règles anti-fraudes, les risques et les activités transactionnelles. Les auteurs présentent le processus semi-automatique de détection de fraudes, mais n'expliquent pas comment l'ontologie est construite, ni comment le peuplement de l'ontologie est réalisé. (Ahmed *et al*., 2021) proposent un algorithme de génération d'alertes basé sur des règles classées par sévérité. Dans ce travail, les auteurs créent l'ontologie *Financial Fraud Detection* (FFD), ils développent des règles de fraudes anti-blanchiment d'argent et un algorithme de génération d'alertes. (Hussaini *et al*., 2022) développent une ontologie pour définir les fraudes financières avec l'objectif d'identifier des patterns pour la prévention et la détection de fraudes. Cette approche est composée de plusieurs étapes : la définition des types de fraudes, la modélisation du comportement des utilisateurs en se basant sur des ontologies existantes. La validation de cette approche est réalisée sur plusieurs jeux de données.

La détection de fraude ne doit pas dégrader la relation avec le client en bloquant les transactions de clients légitimes. Pour cela, le processus de détection de fraude doit également prendre en compte le client en intégrant son profil et ses usages. A notre connaissance, il n'existe pas de travaux sur les transactions du réseau SWIFT, et sur la modélisation des différents acteurs (clients ou agents). Le développement d'une approche basée sur des systèmes de règles avec une ontologie du domaine prenant en compte le profil client et les spécificités du réseau SWIFT pourrait aider les institutions financières à améliorer leur système de lutte contre la fraude.

3. Approche

Nous présentons dans cette section l'ontologie du domaine développée dans le cadre de nos travaux. Nous construisons dans un premier temps l'ontologie du domaine (concepts et propriétés). Cette dernière est peuplée à partir de transactions SWIFT. Nous définissons à l'aide d'experts des règles de fraudes basées sur les transactions et les clients. Cela nous permet d'étiqueter et de détecter des transactions frauduleuses de l'ontologie peuplée. Ensuite, nous exécutons des requêtes pour identifier les transactions frauduleuses et les présenter aux experts sous forme structurée à l'aide d'un outil de visualisation basé sur les graphes.

Nous avons sélectionné avec l'aide d'experts du domaine les informations les plus pertinentes pour la détection de fraude. Les informations concernant les clients ont été sélectionnées de la base de connaissances KYC. Dans l'ontologie proposée, les concepts présentés dans le tableau 1 sont divisés en trois concepts principaux : le **client**, l'institution financière qui est un **agent** débiteur, créditeur ou intermédiaire et la **transaction**. Les propriétés de données et d'objet ont également été définis à l'aide des experts.

TABLEAU 1. *Concepts*

Name	**Description**
Activity	Le domaine d'activité
Agent	Institution financière (banque du client, banque de l'intermédiaire)
Agent:CreditorAgent	Institution financière attachée au client créditeur
Agent:DebtorAgent	Institution financière attachée au client débiteur
Agent:IntermediaryAgent	Institution financière intermédiaire entre deux institutions financières
Customer	Client créditeur et débiteur (n'est pas une institution financière)
Customer:Organization	Responsable de vente de biens et services
Customer:Organization:Association	Organisations, associations et ONG. Pas utilité à cause d'ambiguïté juridique.
Customer:Organisation:Company	Entreprise publique ou privée
Customer:Person	Personne
Transaction	Transfert d'argent entre deux clients

Nous présentons dans le tableau 2 quelques règles d'inférences avec le langage SWRL que nous avons défini avec les experts.

TABLEAU 2. *Régles*

Nom	**Description**
Règle 1 : PaysEtranger	La transaction est effectuée dans un pays différent de celui de l'agent
Règle 2 : ZoneUE-US	le client se situe en Europe ou aux États-Unis et il réalise une transaction avec une devise différente de l'euro ou du dollar.
Règle 3 : MontantJour	Examine le montant total des transactions effectuées
Règle 4 : NbDevises	Transactions dans plus de 5 devises
Règle 5 : Expiration	Vérifie la date et le montant
Règle 6 : Triangulaire	Vérifie le transfert d'argent

1: $\textbf{Transaction}(?t) \wedge \textbf{TransactionCountry}(?t, ?tc) \wedge \textbf{Agent}(?a) \wedge \textbf{AgentCountry}(?a, ?ac) \wedge \textbf{notEqual}(?tc, ?ac) \rightarrow \textbf{IsFraud}(?t, true)$

2: $\textbf{Transaction}(?t) \wedge \textbf{currency}(?t, ?c) \wedge \textbf{TransactionCountry}(?t, ?tc) \wedge (\textbf{notEqual}(?c, "euro") \vee \textbf{notEqual}(?c, "dollar")) \wedge (\textbf{equal}(?tc, "Fr") \vee \textbf{equal}(?tc, "Usa")) \rightarrow \textbf{IsFraud}(?t, true)$

3: $\textbf{Transaction}(?t) \wedge \textbf{amount}(?t, ?amt) \wedge \textbf{Agent}(?a) \wedge \textbf{DailyAmount}(?a, ?dam) \wedge \textbf{lessThan}(?amt, 100) \wedge \textbf{greaterThan}(?dam, 10000) \rightarrow \textbf{IsFraud}(?t, true)$

4: $\textbf{Transaction}(?t) \wedge \textbf{Agent}(?a) \wedge \textbf{AgentCurrencyCount}(?a, ?count) \wedge \textbf{greaterThan}(?count, 5) \rightarrow \textbf{IsFraud}(?t, true)$

5: Transaction$(?t) \wedge$ **Agent**$(?a) \wedge$ **LastTransactionDate**$(?ltd, ?a) \wedge$
greaterThan$(monthsBetween(now(), ?ltd), 6) \wedge$ **amount**$(?t, ?amt) \wedge$
greaterThan$(?amt, 1000) \rightarrow$ **IsFraud**$(?t, true)$

6: Transaction$(\texttt{?t1}) \wedge$ **TransactionDate**$(\texttt{?t1}, \texttt{?td1}) \wedge$ **DebitorAgent**$(\texttt{?da}) \wedge$
hasTransactionWithSameCreditorAndDate$(\texttt{?da}, \texttt{?td1}, \texttt{?t2List}) \wedge$
$\texttt{ListSize}(\texttt{?t2List}, \texttt{?size}) \wedge$ **GreaterThan**$(\texttt{?size}, 0) \wedge$ **CreditorAgent**$(\texttt{?t2}, \texttt{?ca}) \wedge$
TransactionDate$(\texttt{?t2}, \texttt{?td2}) \wedge$ **CreditorAgent**$(\texttt{?ca}) \wedge \texttt{Equals}(\texttt{?ca}, \texttt{?da}) \wedge$
Equals$(\texttt{?td2}, \texttt{?td1}) \wedge$ **Amount**$(\texttt{?t1}, \texttt{?amount1}) \wedge$ **ListContains**$(\texttt{?t2List}, \texttt{?t2}) \wedge$
Amount$(\texttt{?t2}, \texttt{?amount2}) \wedge$ **Equals**$(\texttt{?amount1}, \texttt{?amount2}) \rightarrow$ **IsFraud**$(\texttt{?t1}, \texttt{true})$

Les expérimentations ont été réalisées avec un jeu de données de 1000000 de transactions provenant du réseau SWIFT. Il s'agit de transactions SWIFT où nous avons extrait les champs relatifs à la transaction et au client. Avant d'utiliser notre ontologie, il était important de s'assurer que l'ontologie était valide et cohérente. Pour ce faire, nous avons soumis notre ontologie à des experts financiers qui l'ont évaluée en termes de vocabulaire, de concepts, de hiérarchie des données et de sémantique. L'ontologie est créée avec protégé 5.5.0 [2]. Nous avons également utilisé les raisonneurs Fact++ et HermiT pour vérifier la consistance et la cohérence de notre ontologie financière. Les raisonneurs ont été utilisés pour vérifier que les données ont été correctement liées entre elles et que les inférences logiques sont cohérentes avec les données dans l'ontologie. Le peuplement d'ontologie financière est une étape cruciale pour analyser les données financières et identifier les transactions frauduleuses et suspectes. Il consiste à créer des instances, des concepts, des propriétés d'objet et des propriétés de données d'ontologie. Avec *Owlready2* [3], nous avons peuplé notre ontologie avec 1000000 transactions financières. Après avoir peuplé l'ontologie, nous avons utilisé *SPARQL* pour exécuter des requêtes pour identifier les transactions suspectes ou frauduleuses. Finalement, pour faciliter l'analyse des données et visualiser les relations entre les transactions, nous avons utilisé la bibliothèque *Pyvis* [4] pour afficher les résultats de nos requêtes *SPARQL* sous forme de graphe (figure 1).

4. Conclusion

Nous avons proposé dans cet article une ontologie de domaine pour les transactions interbancaires du réseau SWIFT avec les informations du client (KYC). Notre travail s'inscrit dans le cadre d'un projet de collaboration avec l'entreprise SKAIZen Group. La construction de l'ontologie OntoFiC permet la modélisation des transactions SWIFT et des clients. Cette ontologie est peuplée à partir des transactions issues d'un jeu de données réel. Notre approche de détection de fraudes repose sur le raisonnement basé sur les règles, nous proposons également de visualiser nos requêtes SPARQL avec un outil de visualisation et la bibliothèque *pyvis*. La validation de notre approche a été réalisée à travers des expérimentations, nous avons obtenu des résultats

2. https://protege.stanford.edu
3. http://owlready2.readthedocs.io
4. https://pyvis.readthedocs.io

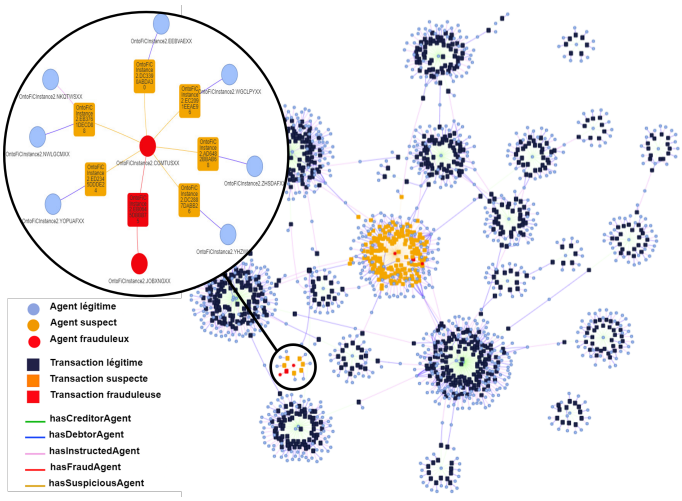

FIGURE 1. *Visualisation des transactions frauduleuses*

prometteurs. Cette approche peut être utilisée en complément d'outils de prévention et de détection de fraudes pour les institutions financières. Dans nos travaux futurs, nous prévoyons d'étendre notre ontologie, en générant dynamiquement de nouvelles règles de détection de transactions frauduleuses avec les techniques d'apprentissage automatique.

Remerciements Ce travail est soutenu à la fois par l'entreprise SKAIZen Group, l'ANRT et l'ANR (France Relance).

Bibliographie

Ahmed M., Ansar K., Muckley C. B., Khan A., Anjum A., Talha M. (2021). A semantic rule based digital fraud detection. *PeerJ Computer Science*.

Attigeri G., M M M., Pai R., Kulkarni R. (2018, 01). Knowledge base ontology building for fraud detection using topic modeling. *Procedia Computer Science*, vol. 135, p. 369-376.

Auger B., Chergui H., Chehade Y., Kadri J. E., Abrouk L., Cabioch N. (2022). Construction d'une ontologie dans le domaine financier pour la détection de fraudes. In *Inforsid*, p. 157–162.

El Orche A., Bahaj M., Ain Alhayat S. (2018). Ontology based on electronic payment fraud prevention. *Faculty of Sciences Technologies HASSAN 1*.

Hussaini A., Guessoum Z., Laurent E. (2022, 09). Elaboration of financial fraud ontology. In, p. 277-285.

Jabbari A., Sauvage O., Zeine H., Chergui H. (2020). A french corpus and annotation schema for named entity recognition and relation extraction of financial news. In *Proceedings of the 12th language resources and evaluation conference*, p. 2293–2299.