Prédiction du risque de crédit bancaire sensible aux coûts financiers en intégrant des descripteurs extraits des graphes

Victor Nico DJIEMBOU TIENTCHEU[1] **and Armel Jacques NZEKON NZEKO'O**[1]

[1]Université de Yaoundé I, Cameroun

*E-mail : Victor nico.djiembou@facsciences-uy1.cm, Armel armel.nzekon@facsciences-uy1.cm

---

Résumé

Les prêts sont des opérations financières très importantes pour le développement et la croissance économique d'un pays, car ces derniers facilitent la création et la croissance des entreprises et donc l'emploi de plus de personnes tant par les entreprises privées, publiques ou parapubliques.

Les non remboursements des prêts ont des coûts importants sur les institutions financières prêteuses, pouvant entraîner leur faillite et donc détruire tout le système de prêt et constituer par là, un frein au développement économique. Il est donc nécessaire de pouvoir prédire efficacement si un prêt sera remboursé ou non par l'emprunteur.

A cet effet, la question de prédiction du risque de crédit est devenu un domaine majeur dans lequel des chercheurs en Intelligence artificielle proposent des modèles qui prédisent la classe d'un prêt à partir des attributs standards qui le décrivent dans l'institution préteuse. Ces attributs standards n'étant pas suffisants pour avoir les meilleures prédictions, ces dernières années, plusieurs travaux portent sur la création de nouveaux attributs descriptifs à mettre en entrée des modèles classiques de prédiction dans le but d'améliorer leur performance.

C'est le cas des récents travaux sur l'extraction de nouveaux descripteurs des prêts modélisés par un graphe multicouches dans lesquels une seule application du PageRank personnalisé sur le graphe multicouches permet d'extraire les nouveaux descripteurs des différents prêts considérés.

Les travaux actuels sur les graphes multicouches ont pour limites de ne pas être suffisamment personnalisés par prêt, de ne pas considérer les classes des prêts dans le processus de construction du graphe lors de l'apprentissage, de ne pas proposer de stratégie pour le choix des attributs à considérer comme couches du graphe construit et enfin de ne pas évaluer leur impact sur les coût financiers qui sont un aspect important pour les institution prêteuses.

Dans ce mémoire, nous proposons d'intégrer les classes des prêts dans le processus de construction des graphes multicouches, et d'appliquer le PageRank personnalisé par prêt pour extraire les nouveaux descripteurs des ces graphes. Par ailleurs, nous proposons un protocole de sélection des attributs à considérer comme couches du graphe multicouches, et effectuons une évaluation des coûts financiers des modèles de prédiction du risque de crédit construits à partir des données enrichies par les nouveaux descripteurs.

Des expérimentations sont menées sur 04 jeux de données, en considérant 06 modèles classiques de prédiction du risque de crédit (LDA, SVM, LR, DT, RF, XGBoost) et 03 métriques d'évaluation des performances des modèles (Accuracy, F1-score, Cost), dont l'une sensible aux coûts financiers. Des valeurs de SHapley sont considérés pour évaluer l'importance des nouveaux descripteurs.

Nous observons que notre approche permet d'améliorer les meilleurs modèles de l'existant dans C% des cas, de plus le fait d'intégrer les informations de classe permet de garantir une réductions

des coûts financiers de plus de Y% dans la majorité de cas.

Mots-Clés
risque de crédit, sciences des réseaux, graphe multicouches

I INTRODUCTION

1.1 Contexte d'application et d'intérêt

Les prêts financiers sont une opération importante dans la croissance économique dans le monde car elle sont utilisé pour subventionner les projets organismes gouvernementaux et des particuliers.

Les projets d'urbanisation et des recherches les plus poussées dans le monde n'existe que parce que les prêts financiers existent.

1.2 Transition vers la problématique

Crise de subprime en 2008

Machine learning pour la prédiction du risque de crédit financiers

manque de données et caractéristiques descriptives

Création de nouvelle dimension descriptives dans les jeux de données en utilisant des graphes multicouches

La logique de PageRank personnalisation ne se porte pas sur chaque information emprunteurs mais sur celui du réseau formé

Il existe pas un mécanismes pour identifier de façon exacts les attributs l'ensembles de relations les plus pertinentes à analyser

La modélisation ne prend pas en compte les informations de décisions de ses historiques pourtant connu

L'approche ne met pas un intérêt sur l'impact de la solution en terme coûts financiers pour l'entité prêteuses.

1.3 Problème

Il est question pour nous dans ce memoire de trouver comment proposer à la fois une façon d'améliorer la personnalisation du PageRank, prendre en compte la décision de prêt dans la modélisation graphe biparti multicouches, de sélectionner les attributs descriptives optimales pour la construction du graphe et mettre ce pieds un métrique de

1.4 Objectif

proposer un PageRank personnalisation porté sur un seul emprunteur à la fois

proposer un protocole qui va permettre d'identifier les relations les plus pertinentes à analyser

proposer une cadre de modélisation graphe biparti multicouches qui incorpore les informations de décision des prêts historique.

1.5 Contribution

proposer une meilleur façons d'extraire des descripteurs des graphes multicouches en proposant

- un PageRank personnalisation porté sur un seul emprunteur à la fois
- un protocole qui va permettre d'identifier les relations les plus pertinentes à analyser
- une cadre de modélisation graphe biparti multicouches qui incorpore les informations de décision des prêts historique.
- une métrique pour évaluer les coûts financiers.

1.6 Plan du mémoire

Le reste de ce memoire se présentera comme suit, dans la section suivante nous présenterons l'état de l'art sur la prédiction du risque de crédit financier, dans la troisième section. Nous présentons notre solution, un cadre d'extraction optimale de descripteurs dans des graphes biparti multicouches. La quatrième section présente notre cadre expérimentale. Et enfin, en section 5 nous conclurons notre travail.

II PRÉDICTION DU RISQUE DE CRÉDIT

l'évaluation du risque de crédit par des modèles de machine learning classique + manque de données + non représentativité + non equivalence des coûts

augmentation de donnée avec les méthodes de sampling (over et under)

méthodes de création de nouvelles attributs descriptives

- graphes complet
- graphe multivue
- graphe biparti
- graphe biparti multicouche

III PRISE EN COMPTE DES DÉCISIONS DANS LA MODÉLISATION GRAPHE BIPARTI MULTICOUCHES ET PERSONNALISATION DU PAGERANK À UN SEUL EMPRUNTEUR POUR UNE PRÉDICTION DU RISQUE DE CRÉDIT SENSIBLE AUX COÛTS FINANCIERS

graphe biparti multicouches

PageRank personnalisé

modèles de machine learning

PageRank Personnalisé à un emprunt

Graphe biparti multicouches intégrant les informations de décision

Protocole de selection des k meilleurs attributs devant servir à la construction du graphes biparti multicouches à k couches.

IV EXPÉRIMENTATIONS

4.1 Description du jeux de données

AFB, CREDIT RISK, GERMAN, JAPAN (nombre de ligne (exemple), nombre de colonnes, nombre de colonnes numériques, nombre d'attributs catégoriel, nombre d'exemples positif, nombre d'exemples négatifs)

4.2 Évaluation et paramétrage de modèles

Acc + F1 + Cost

Approches
- MlC
- MCA

Logiques
- GLO
- PER
- GAP

modeles
- MX
- CX
- CY
- CXY

4.3 Résultats

SHAP + Tableaux

V CONCLUSION

5.1 rappel du problème abordé

Il etait question pour nous dans ce memoire de trouver comment proposer à la fois une façon d'améliorer la personnalisation du PageRank, prendre en compte la décision de prêt dans la modélisation graphe biparti multicouches, de sélectionner les attributs descriptives optimales pour la construction du graphe et mettre ce pieds un métrique de

5.2 Idée de solution

Prise en compte des décisions dans la modélisation graphe biparti multicouches et personnalisation du PageRank à un seul emprunteur pour une prédiction du risque de crédit sensible aux coûts financiers

5.3 Démarche

proposer un PageRank personnalisé porté sur un seul emprunteur à la fois

proposer un protocole qui va permettre d'identifier les relations les plus pertinentes à analyser

proposer une cadre de modélisation graphe biparti multicouches qui incorpore les informations de décision des prêts historique.

une métrique de coûts financiers

5.4 Les principaux résultats

5.5 Les perspectives

prendre en compte les données numérique dans la modélisation

proposer de nouvelles stratégies d'exploitation du graphes autres que le PageRank personnalisé.

RÉFÉRENCES

A ANNEXE 1

Dans Bibtex, comment écrire une citation d'un article de ARIMA (example : **arima**) sans rien oublier et dans le bon format ? Voir la structure dans les commentaires à la fin de *arima.tex*.

B REMERCIEMENTS

Nous tenons à remercier tous nos partenaires financiers : ANR ..., ERC ..., agences de financement, ...

C BIOGRAPHIE

Il est possible ici d'insérer de courtes biographies des auteurs.