

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

Victor DJIEMBOU¹ and Armel NZEKON¹

¹Université de Yaoundé I, Cameroun

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

Abstract

Le prêt est l'une des principales sources d'enrichissement des banques, mais est également une source de perte financière. Pour optimiser leurs profits, il est donc nécessaire pour les banques de prédire si un emprunteur remboursera ou pas son crédit. Afin d'aider les banques à aborder efficacement ce challenge, plusieurs techniques d'apprentissage automatique ont été proposées durant cette dernière décennie pour la question spécifique de prédiction du risque de crédit bancaire. Quelques unes de ces techniques reposent sur des graphes

L'amélioration de la capacité de ses institutions à évaluer le risque associé à une demande de prêt a été confronté à des difficultés parmi lesquelles la faible présence de données pour entraîner les modèles d'AA et la faible représentativité dans les jeux de données existants. L'analyse des comportements des emprunteurs et des liens avec leur entourage au moyen de modèles graphes a été longtemps utiliser pour augmenter les jeux de données existant avec des informations pertinentes pour combler le manque de données et relativement leur représentativité.

Ainsi de nombreux travaux en IA ont réussi jusqu'ici à proposer un moyen d'analyser les emprunteurs suivant deux dimensions. Différentes paires d'attributs catégorielles sont prise de façon aléatoire dans l'ensemble d'informations et sont modélisé au moyen de graphes multicouches biparti où chaque couche met les emprunteurs en relations avec la valeur d'attribut associée constituant cette couche. Ensuite des informations pertinentes comme le PageRank avec une personnalisation inter-couche, intra-couche et collective de ces graphes sont calculées via l'analyse de graphes. Néanmoins, l'aléa dans le choix d'attributs du jeu données, la limite du nombre de dimension et encore la limitation au attributs catégorielles constituent une limite car de façon hypothétique, je suis en relation avec un ensemble d'individu parce que je partage au moins une information avec cette ce groupe. Alors, pour résoudre ses limites, nous adaptions d'abord l'algorithme de PageRank personnalisé multicouche pour prendre en compte unique les informations propres à un seul emprunteurs lors de la recherche d'information pertinente, ensuite nous proposons un protocole qui permet de choisir les k attributs à utiliser dans un graphe multicouche biparti à k couches, enfin les informations extraites sont intégrées dans des modèles AA notamment afin d'améliorer les métriques de classification de ses modèles. Nous l'avons approuvé à 5 jeux de données benchmark dans le domaine du credit scoring et avons observé que nous améliorons les résultats présentent dans la littérature.

Keywords

risque de crédit, graphe, graphe multicouche

I INTRODUCTION

La science des réseaux s'est imposée comme une théorie fondamentale et un outil essentiel pour comprendre, modéliser et analyser les réseaux de communication. La science des réseaux s'est imposée comme une théorie fondamentale et un outil essentiel pour comprendre, décrire, modéliser et analyser les systèmes complexes d'entités en interaction qui apparaissent naturellement, par exemple, en biologie, en sociologie, en finance et en économie, parmi beaucoup d'autres [4]. Dans la vie réelle, les objets sont souvent liés par plus d'un type de relation. Par exemple, dans les réseaux sociaux en ligne, deux personnes peuvent être amies sur Facebook et se suivre sur Twitter. Ces réseaux sont généralement appelés réseaux multicouches, dont la théorie et le développement de modèles ont été au cœur de la science des réseaux au cours de la dernière décennie.

Dans le secteur bancaire, on soupçonne depuis longtemps l'existence de défaillances corrélées. Cependant, toutes les études qui ont porté sur la propagation du risque à travers les réseaux se sont concentrées sur la contagion. Cependant, toutes les études portant sur la propagation du risque à travers les réseaux se sont concentrées sur la contagion entre les institutions financières [8,20] ou sur la corrélation des actifs avec l'ensemble de l'économie, proposée pour la première fois dans les accords de Bâle [5]. Plusieurs auteurs ont souligné que ces effets de corrélation affectent les mesures actuelles du risque de crédit [21,49] au niveau individuel, un fait qui a été négligé jusqu'à présent. Compte tenu de l'importance des prêts aux particuliers pour l'ensemble de l'économie et des progrès actuels de la science des réseaux, nous combinons le risque de crédit avec la science des réseaux multicouches pour présenter une méthodologie permettant de calculer explicitement l'impact de la corrélation des défaillances chez les emprunteurs connectés au niveau un à un. Notre proposition vise donc non seulement à mieux comprendre la propagation des défaillances, mais aussi à améliorer les systèmes de notation des demandes, couramment utilisés pour prendre des décisions en matière d'octroi de prêts.

Ce document comporte trois contributions principales. Premièrement, nous présentons un cadre pour créer un réseau bipartite multicouche à partir d'un ensemble de données tabulaires avec deux ou plusieurs variables de connexion. Ces variables ne sont pas nécessairement des variables de réseau explicites, mais peuvent être une propriété partagée ou une caractéristique. Ensuite, nous développons une nouvelle approche pour la prédition du risque de crédit avec une mesure de centralité PageRank multicouche personnalisée qui classe les nœuds du réseau multicouche par rapport à un ensemble de nœuds sources qui peuvent être utilisés pour comprendre le risque corrélé et également pour améliorer les systèmes d'évaluation du crédit. Enfin, nous montrons que les scores d'influence des réseaux multicouches contribuent à l'amélioration des modèles d'évaluation du crédit. Bien que nos résultats soient généraux et applicables à n'importe quel ensemble de données de prêt (et à n'importe quel problème dans lequel des effets corrélés sont suspectés), nous motivons le travail dans un environnement dans lequel le risque de crédit est le plus élevé. motivons notre travail dans un environnement où la corrélation des défaillances est évidente. Nous concentrons notre attention sur les réseaux multicouches dans les prêts agricoles et présentons des algorithmes généraux pour les inclure dans tout modèle prédictif où ils apparaissent. Si les réseaux multicouches ont été étudiés dans le secteur bancaire, jusqu'à présent - et à notre connaissance - leurs effets sur les prêts de détail (prêts à la consommation traditionnels et prêts de gré à gré aux petites et moyennes entreprises) n'ont pas fait l'objet du même niveau d'attention. Cela n'est pas surprenant compte tenu de sa structure : les prêts aux particuliers sont beaucoup plus nombreux, de sorte que toute interaction entre les réseaux nécessite une analyse complexe qui n'est devenue réalisable qu'au cours

de la dernière décennie. Cela dit, il est également naturel de s’attendre à ce que les emprunteurs soient connectés de nombreuses manières complexes, de sorte que l’analyse utilisée pour mesurer leur comportement devrait également suivre ce niveau de complexité.

II SECTION

2.1 Sub-section 1

Pellentesque dignissim ultrices fringilla. Vivamus eu luctus ante, vel bibendum magna. Curabitur elit purus, tincidunt non dui vitae, elementum bibendum neque. Curabitur ullamcorper sit amet justo at hendrerit. Fusce ut arcu imperdiet nibh mollis tempus a aliquet tellus. Quisque pharetra cursus nisi, vel lobortis ante consectetur et. Vivamus sed congue neque. Proin pellentesque risus nec dui consequat rutrum. Vestibulum nunc diam, placerat quis auctor vel, faucibus non justo. Etiam dictum purus neque. Phasellus imperdiet mauris ligula, eu laoreet nisi elementum ut. Sed sed porta massa. Aenean faucibus risus ultrices ornare porta. Quisque faucibus ante a tincidunt vestibulum. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

2.2 Sub-section 2

2.2.1 Sub-sub-section

Suspendisse vel dui nec felis molestie tincidunt. Vestibulum rutrum ligula lacus, ac molestie nulla fermentum ornare. Nulla non nunc euismod, porta lacus vestibulum, malesuada massa. Curabitur massa eros, rutrum sed lectus sed, volutpat semper metus. Mauris hendrerit aliquam commodo. Vivamus fermentum tempus pellentesque. Maecenas a hendrerit urna. In elit ipsum, ultrices non dolor in, pulvinar porttitor lacus. Nunc euismod nibh quis odio condimentum, a feugiat massa rutrum. Nulla erat erat, adipiscing vitae lectus id, consectetur fermentum elit. Nunc eu est eu neque dapibus semper. Nam commodo urna dapibus, tincidunt turpis a, cursus sem. Vivamus venenatis adipiscing mollis. Cras fringilla sodales lobortis. Aliquam aliquet felis id est cursus auctor. Duis sodales tellus vulputate lectus egestas volutpat.

III TABLES AND FIGURES

3.1 Tables

Cf. Table 1.

	S-Length	S-Width	P-Length	P-Width
Setosa	5.006	3.428	1.462	0.246
Versicolor	5.936	2.77	4.26	1.326
Virginica	6.588	2.974	5.552	2.026

Table 1: Morbi malesuada diam at magna condimentum.

3.2 Figures

If possible, bring figures together in a block. Use images at least 300 dpi. Check that caption are legible and in english. See this example : Figure 1 (from <http://www.texample.net/tikz/examples/pgfplots/>). The title must be ended with a dot.

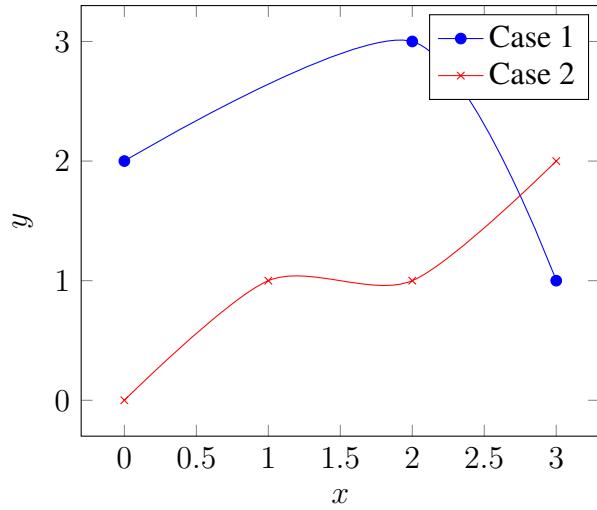


Figure 1: figure de Christian Feuersänger; source: Pgfpplots.



Figure 2: Cameroon Lion (photo credit: Fist name Surname).

IV DEFINITIONS, ALGORITHMS AND FORMULAS

4.1 Definitions

For formulas, use the package `amsthm` and the style `arima` for reliable formatting.

Definition 1: alpha

Curabitur ullamcorper sit amet justo at hendrerit.

Then we set the following definition.

Definition 2:

Etiam sed nulla viverra, ultrices ligula ac, consectetur libero.

For a list, use "bullets" or em dash:

- Nunc id justo scelerisque ;
- metus id enim iaculis tristique.

4.2 Formulas

Formulas Example:

$$Y = M \cdot {}^t M - \beta \langle M \rangle_l \quad (1)$$

where $\langle M \rangle_l$ is the mean vector of a M . β enables the rate of close neighbours to be regulated.

Other formulas example:

$$K * N_c = Cst \pm 0.001\% \quad (2)$$

4.3 Algorithms

For algorithms, use appearance of Algorithm IV.1. If necessary, see the package `listings`.

Algorithm IV.1 Partitioning function of a sorting algorithm.

```
1  partition (array , left , right )
2      pivotIndex := choose-pivot( array , left , right )
3      pivotValue := array [ pivotIndex ]
4      swap array [ pivotIndex ] and array [ right ]
5      storeIndex := left
6      for i from left to right - 1
7          if array [ i ] < pivotValue
8              swap array [ i ] and array [ storeIndex ]
9              storeIndex := storeIndex + 1
10     swap array [ storeIndex ] and array [ right ] // Move pivot to its final place
11     return storeIndex
```

V CONCLUSION AND REFERENCES

5.1 Discussion

Nam id eros massa. Fusce luctus purus a augue ullamcorper, sit amet vehicula mauris tristique. Suspendisse eget pulvinar odio, nec bibendum turpis. Nullam quis lectus porttitor, ullamcorper nisi et, condimentum leo. Quisque sed orci fermentum, rutrum velit eget, ultricies augue. Nunc porttitor consectetur tincidunt. Nulla tincidunt justo enim, vitae dignissim erat mattis ut. Nulla.

5.2 Conclusion

Maecenas egestas metus id enim iaculis tristique. Etiam sed nulla viverra, ultrices ligula ac, consectetur libero. Nullam vitae massa ac odio pharetra condimentum. Maecenas in elementum libero, non gravida quam. Praesent adipiscing consectetur. Vivamus at orci sed augue varius hendrerit. Donec neque metus, dignissim nec erat at, ultricies consequat libero. Donec eget eleifend leo. Aliquam at nunc porta, mollis sapien eu, eleifend tortor. Nam egestas, metus ac pellentesque feugiat, lectus purus ornare est, vitae cursus felis turpis sit amet lacus. Donec consequat massa mi, ac suscipit arcu posuere et. Vivamus et semper risus. Sed ut arcu quam.

REFERENCES

A ANNEX 1

In Bibtex, how to write a citation from an ARIMA article (example: **arima**) without forgetting anything and in the right format? See the structure in comments at the end of *arima.tex*.

B ACKNOWLEDGEMENTS

All our funding partners are gratefully acknowledged: ANR ..., ERC ..., funding agencies, ...

C BIOGRAPHY

Short biographies of the authors can be inserted here.