

Prédiction du risque de crédit bancaire sensible aux coûts financiers avec des descripteurs issus des graphes

Spécification et justification du sujet de recherche

Supervisé par :
Pr. ATSA ETOUNDI ROGER



Faculté des Sciences

Nom :	DJIEMBOU TIENTCHEU VICTOR NICO
Devoir :	1
Reçu le :	2023.12.01
Auteur :	Perçu de Pr. ATSA ETOUNDI ROGER
Mots clés :	Méthodologie, Recherche, Graphe, Finance, Crédit
UE :	INF5019
Matricule :	17T2051
Département :	Informatique
Niveau - Option :	M2 - Sciences des Données

Table des matières

Acronyms	3
1 Contexte	4
2 Problème	4
3 motivations	4
4 Les domaines de recherche : lequel supporte ton sujet ?	4
5 Type de recherche ou courants de pensée qu'on fait dans le domaine du Machine Learning (ML) : Lequel utilises-tu dans ton sujet ?	5
6 Méthodologies de recherche utilisées dans le ML : Laquelle utilises-tu dans ton sujet ? Et comment fonctionnel t-elle ?	6
7 Bases ou socles scientifiques dans le domaine du sujet	7
8 Les dieux ou patrons ou grand chercheurs ou pionniers mondiaux et leurs bibles dans le ML	8
9 Les conférences et journaux de référence des dieux dans le ML	9
10 Axes de recherche dans le ML	10
11 Tâches pour arriver au choix de l'axe de recherche	11
12 Axe de recherche du sujet de recherche	11
13 Conclusion	11
Annexe	12
Références	13

Acronymes

ML	Machine Learning
IA	Intelligence Artificielle
CV	Computer Vision
NLP	Natural Language Processing
HCI	Human-Computer Interaction
MLN	MultiLayer Networks
SVM	Support Vector Machine
JMLR	Journal of Machine Learning Research
ICLR	International Conference on Learning Representations
NeurIPS	Neural Information Processing Systems
AAAI	Association for the Advancement of Artificial Intelligence
TPAMI	Transactions on Pattern Analysis and Machine Intelligence
JAIR	Journal of Artificial Intelligence Research
MLJ	Machine Learning Journal
DL	deep learning

1 Contexte

2 Problème

3 motivations

Il serait intéressant pour nous étudiants de pouvoir à partir de cours de pouvoir expertiser sur

- comment penser et mettre sur pieds un projet avec logique de parallélisation dans les langages c et c++ ?
- quels sont les bibliothèques utilisées dans ses distincts langages pour paralléliser le code ?
- quels sont les contraintes d'implémentation dans chaque langage ?

4 Les domaines de recherche : lequel supporte ton sujet ?

De la lecture faite, une liste probablement non exhaustive de 10 grand domaines de recherche en informatique a pu être identifié parmi lesquels :

1. **Intelligence Artificielle (IA)** : L'IA concerne la conception de systèmes informatiques capables de simuler des comportements intelligents, tels que la perception sensorielle, le raisonnement, l'apprentissage et la prise de décision.
2. **Apprentissage automatique (ML)** : Le domaine de l'apprentissage automatique se concentre sur le développement d'algorithmes et de modèles qui permettent aux machines d'apprendre à partir de données et d'améliorer leurs performances sans être explicitement programmées.
3. **Vision par ordinateur (Computer Vision (CV))** : Ce domaine vise à permettre aux ordinateurs de comprendre, analyser, interpréter et reconnaître des images et des vidéos, en simulant la vision humaine.
4. **Traitement du langage naturel (Natural Language Processing (NLP))** : Le NLP concerne le traitement automatique du langage humain, y compris la compréhension, la génération, la traduction et l'analyse de texte ou de parole.
5. **Réseaux neuronaux et architectures de calcul neuromorphique** : Ce domaine se concentre sur la modélisation et la simulation de réseaux neuronaux artificiels, qui s'inspirent du fonctionnement du cerveau humain, ainsi que sur le développement de matériel informatique spécialisé pour des calculs neuromorphiques.
6. **Informatique quantique** : L'informatique quantique explore les principes de la physique quantique pour développer de nouvelles formes de calcul et de traitement de l'information, qui pourraient potentiellement résoudre des problèmes complexes de manière plus efficace que les ordinateurs classiques.
7. **Systèmes distribués et parallèles** : Ce domaine concerne la conception, le développement et l'optimisation de systèmes informatiques qui utilisent plusieurs ressources de calcul en parallèle, tels que des clusters, des grappes de serveurs ou des réseaux de machines interconnectées.
8. **Réseaux et sécurité informatique** : Ce domaine englobe la conception, le déploiement et la sécurisation des réseaux informatiques, ainsi que la protection des données, la détection des intrusions et la résolution des problèmes de sécurité.
9. **Interaction homme-machine (Human-Computer Interaction (HCI))** : L'HCI se concentre sur la conception et l'évaluation des interfaces utilisateur pour améliorer l'interaction entre les humains et les machines, en tenant compte des aspects ergonomiques, cognitifs, émotionnels et sociaux.

10. **Informatique théorique et algorithmique** : Ce domaine explore les fondements mathématiques de l'informatique, y compris la conception et l'analyse d'algorithmes, la théorie de la complexité, la théorie des graphes, la logique formelle et la théorie de l'information.

La prédiction du risque de crédit vise à trouver un algorithme donnant les moyen à une machine d'apprendre à partir des données comment estimer si un emprunteur pourra ou pas rembourser son emprunt avec une certaine confiance. Ceci nous plonge au coeur du domaine du ML.

5 Type de recherche ou courants de pensée qu'on fait dans le domaine du ML : Lequel utilises-tu dans ton sujet ?

Le domaine du ML se veut de posséder une multitude de type de recherche parmi lesquelles :

1. **Recherche fondamentale** : Ce type de recherche se concentre sur la compréhension théorique des algorithmes et des modèles d'apprentissage automatique. Elle vise à élucider les principes sous-jacents de l'apprentissage automatique, à établir des garanties mathématiques de performance et à explorer les limites et les capacités des modèles.
2. **Recherche appliquée** : La recherche appliquée en apprentissage automatique se concentre sur l'application des techniques d'apprentissage automatique pour résoudre des problèmes spécifiques dans des domaines tels que la vision par ordinateur, le traitement du langage naturel, la robotique, la bioinformatique, la finance, etc. Elle vise à développer des modèles et des algorithmes adaptés aux tâches spécifiques et à les mettre en œuvre dans des applications concrètes.
3. **Recherche en optimisation** : L'optimisation est un aspect clé de l'apprentissage automatique, et la recherche dans ce domaine se concentre sur le développement de techniques d'optimisation efficaces pour entraîner des modèles d'apprentissage automatique, sélectionner des hyperparamètres et résoudre d'autres problèmes liés à l'optimisation des performances des modèles.
4. **Recherche en interprétabilité** : L'interprétabilité de l'apprentissage automatique est un domaine de recherche émergent qui vise à comprendre et à expliquer les décisions prises par les modèles d'apprentissage automatique. La recherche dans ce domaine explore des techniques et des méthodes pour rendre les modèles plus transparents et compréhensibles, afin de faciliter leur adoption dans des domaines sensibles tels que la santé, la finance et le droit.
5. **Recherche en apprentissage en ligne (Online Learning)** : L'apprentissage en ligne concerne la mise à jour continue des modèles d'apprentissage automatique à mesure que de nouvelles données deviennent disponibles. La recherche dans ce domaine se concentre sur le développement d'algorithmes qui sont capables de s'adapter et d'apprendre à partir de flux de données en temps réel, tout en maintenant de bonnes performances.
6. **Recherche en apprentissage fédéré (Federated Learning)** : L'apprentissage fédéré est une approche de l'apprentissage automatique qui permet de collaborer et de construire des modèles sans partager les données brutes entre différentes entités. La recherche dans ce domaine se concentre sur le développement de techniques pour entraîner des modèles efficacement tout en préservant la confidentialité et la sécurité des données.
7. **Recherche en apprentissage supervisé** : Ce type de recherche se concentre sur l'apprentissage à partir de données étiquetées. Il vise à développer des algorithmes et des modèles capables de prédire une sortie souhaitée en se basant sur des exemples d'entraînement où les entrées sont associées à des étiquettes connues. C'est le cas des problèmes de classification et de régression.
8. **Recherche en apprentissage non supervisé** : Contrairement à l'apprentissage supervisé, l'apprentissage non supervisé vise à découvrir des structures et des modèles intrinsèques

dans les données non étiquetées. Cela peut inclure des techniques telles que la clustering, la réduction de dimension, la détection d'anomalies, etc.

9. **Recherche en apprentissage par renforcement** : Ce type de recherche concerne l'apprentissage à partir de l'interaction avec un environnement. L'agent d'apprentissage prend des actions dans un environnement et reçoit des récompenses ou des pénalités en fonction de ses actions. L'objectif est de développer des algorithmes qui maximisent la récompense cumulée au fil du temps.

Le présent sujet de recherche s'inscrit dans les cadres suivant :

- **Recherche appliquée** : car le sujet cherche à résoudre un problème spécifique dans le domaine bancaire en appliquant des techniques d'apprentissage automatique pour prédire le risque de crédit. De plus, utilise des descripteurs issus des graphes pour enrichir les modèles de prédiction.
- **Recherche en apprentissage supervisé** : utilise des données historiques étiquetées pour entraîner un modèle de prédiction du risque de crédit. Le but est de prédire si un emprunteur est susceptible de faire défaut ou non en utilisant des descripteurs issus des graphes.
- **Recherche en apprentissage profond avec des graphes** : Étant donné que les descripteurs proviennent de graphes, il est possible d'explorer des techniques d'apprentissage profond, telles que les graphes multi-couches (MultiLayer Networks (MLN)), pour capturer les relations complexes entre les entités financières et améliorer la prédiction du risque de crédit.
- **Recherche en interprétabilité des modèles** : Étant donné l'importance des coûts financiers dans votre sujet, il est possible d'explorer la recherche en interprétabilité des modèles d'apprentissage automatique. Ceci aidera à comprendre comment les descripteurs issus des graphes contribuent aux prédictions du risque de crédit et à fournir des explications compréhensibles pour les décisions prises par le modèle.

6 Méthodologies de recherche utilisées dans le ML : Laquelle utilises-tu dans ton sujet ? Et comment fonctionne-t-elle ?

Ses méthodologies sont :

1. Définition du problème : La première étape consiste à définir clairement le problème de recherche ou le problème de prédiction que l'on souhaite résoudre. Cela implique de définir les objectifs, de spécifier les données d'entrée et de sortie, et de comprendre le contexte et les contraintes du problème.
2. Collecte des données : Une fois le problème défini, il est nécessaire de collecter les données pertinentes pour l'entraînement, la validation et le test du modèle. Cela peut impliquer la collecte de nouvelles données ou l'utilisation de jeux de données existants disponibles publiquement ou privément.
3. Exploration des données : Avant de construire un modèle, il est essentiel d'explorer et d'analyser les données pour comprendre leur distribution, identifier les éventuels problèmes de qualité des données, détecter les valeurs aberrantes et les données manquantes, et effectuer une visualisation des données.
4. Préparation des données : Cette étape implique le nettoyage des données en éliminant les valeurs aberrantes, en traitant les données manquantes, en normalisant les valeurs, en sélectionnant les caractéristiques pertinentes et en divisant les données en ensembles d'entraînement, de validation et de test.
5. Conception du modèle : Une fois les données préparées, la conception du modèle intervient. Cela implique de choisir l'architecture du modèle, de sélectionner les hyperparamètres, d'ini-

tialiser les poids du modèle, et de définir les fonctions de perte (loss functions), les algorithmes d'optimisation et les métriques d'évaluation.

6. Entraînement du modèle : Dans cette étape, le modèle est entraîné sur les données d'entraînement en ajustant les poids du modèle à l'aide d'algorithmes d'optimisation. L'objectif est de minimiser la fonction de perte en utilisant des méthodes telles que la descente de gradient stochastique (SGD) ou ses variantes.
7. Évaluation du modèle : Une fois le modèle entraîné, il est évalué sur les données de validation pour mesurer sa performance et son aptitude à généraliser sur de nouvelles données. Différentes métriques d'évaluation peuvent être utilisées en fonction du problème, comme la précision, le rappel, l'aire sous la courbe ROC, etc.
8. Réglage du modèle : Si la performance du modèle n'est pas satisfaisante, des ajustements peuvent être apportés, tels que la modification des hyperparamètres, la modification de l'architecture du modèle, l'utilisation de techniques de régularisation ou l'ajout de nouvelles fonctionnalités.
9. Test du modèle : Une fois le modèle finalisé, il est testé sur des données indépendantes (ensemble de test) pour évaluer sa performance réelle. Cela permet de s'assurer que le modèle généralise correctement et n'est pas surajusté (overfitting) aux données d'entraînement.
10. Analyse des résultats : Enfin, une analyse approfondie des résultats est effectuée pour comprendre les prédictions du modèle, identifier les erreurs, détecter les biais potentiels et extraire des informations et des connaissances utiles.

7 Bases ou socles scientifiques dans le domaine du sujet

Les socles du ML sont :

1. Statistiques : Les techniques de machine learning sont basées sur des concepts statistiques tels que l'estimation, l'inférence et l'analyse de données. La compréhension des distributions de probabilité, des tests d'hypothèses et des méthodes d'échantillonnage est cruciale pour évaluer les modèles et interpréter les résultats.
2. Calcul différentiel et calcul vectoriel : Le calcul différentiel est utilisé pour l'optimisation des modèles, où des algorithmes d'optimisation itératifs sont appliqués pour ajuster les paramètres des modèles en fonction des données. Le calcul vectoriel est utilisé pour représenter et manipuler les données et les caractéristiques dans un espace multidimensionnel.
3. Algèbre linéaire : L'algèbre linéaire est utilisée pour représenter et manipuler des matrices et des vecteurs, qui sont des structures de données couramment utilisées dans les modèles de machine learning tels que les réseaux de neurones et les méthodes de régression.
4. Théorie des graphes : La théorie des graphes est utilisée pour représenter les relations entre les données, en particulier dans les modèles d'apprentissage non supervisé tels que les réseaux de neurones auto-encodeurs et les méthodes de regroupement (clustering).
5. Apprentissage statistique : Les techniques d'apprentissage statistique, telles que la régression linéaire, la régression logistique, les machines à vecteurs de support (SVM) et les arbres de décision, constituent la base des méthodes de machine learning. La compréhension de ces techniques est essentielle pour construire et évaluer des modèles prédictifs.
6. Théorie de l'information : La théorie de l'information fournit des mesures quantitatives de l'information et de l'entropie qui sont utilisées dans l'apprentissage automatique, par exemple, pour évaluer la performance des modèles de compression ou pour quantifier la redondance des données.
7. Informatique distribuée et parallèle : Avec l'explosion des données, l'informatique distribuée et parallèle est devenue essentielle pour entraîner des modèles de machine learning sur de

grandes quantités de données. La compréhension des techniques de parallélisme, de la gestion des ressources et de la communication entre les nœuds est importante pour l'efficacité et les performances des systèmes de machine learning distribués.

8 Les dieux ou patrons ou grand chercheurs ou pionniers mondiaux et leurs bibles dans le ML

Les pionniers du ML de façon généraliste sont :

1. Arthur Samuel : Considéré comme l'un des pionniers de l'apprentissage automatique, Arthur Samuel a développé le concept de "machine learning" dans les années 1950. Il est connu pour avoir créé un programme d'échecs capable d'apprendre et de s'améliorer à travers l'expérience.
— Article de référence : "Some Studies in Machine Learning Using the Game of Checkers" (1959)
2. Frank Rosenblatt : Frank Rosenblatt est connu pour avoir développé le perceptron, un modèle de réseau de neurones artificiels, dans les années 1950. Son travail a jeté les bases des réseaux de neurones et de l'apprentissage supervisé.
— Article de référence : "The Perceptron : A Probabilistic Model for Information Storage and Organization in the Brain" (1958)
3. Geoffrey Hinton : Considéré comme l'un des pionniers de l'apprentissage profond (deep learning), Geoffrey Hinton a joué un rôle essentiel dans le développement des réseaux de neurones profonds. Ses travaux ont contribué à des avancées significatives dans la vision par ordinateur, la reconnaissance vocale et la traduction automatique.
— Article de référence : "Deep Learning" (2007)
— Article de référence : "A Fast Learning Algorithm for Deep Belief Nets" (2006)
4. Yann LeCun : Yann LeCun est un chercheur français connu pour ses contributions majeures à l'apprentissage profond et à la vision par ordinateur. Il est notamment à l'origine du développement du modèle de réseau de neurones convolutifs (CNN) et a joué un rôle clé dans les avancées de la reconnaissance d'images.
— Article de référence : "Gradient-based Learning Applied to Document Recognition" (1998)
— Article de référence : "Convolutional Networks for Image Recognition" (1998)
5. Vladimir Vapnik : Vladimir Vapnik est un chercheur connu pour son travail sur les machines à vecteurs de support (SVM). Ses contributions ont eu un impact significatif dans le domaine de l'apprentissage supervisé et des problèmes de classification.
— Livre de référence : "The Nature of Statistical Learning Theory" (1995)
— Article de référence : "A Support Vector Method for Multivariate Density Estimation" (1998)
6. Andrew Ng : Andrew Ng est un chercheur et entrepreneur qui a contribué à populariser l'apprentissage automatique. Il a cofondé le programme d'apprentissage automatique de Stanford et a été l'un des cofondateurs de Coursera, une plateforme d'apprentissage en ligne. Il a également été directeur de recherche chez Google Brain.
— Article de référence : "Sparse Autoencoder" (2008)
— Article de référence : "Deep Learning" (2012)
7. Yoshua Bengio : Yoshua Bengio est un chercheur canadien qui a apporté des contributions majeures à l'apprentissage profond et aux réseaux de neurones. Il est considéré comme l'un des pionniers de l'apprentissage profond et ses travaux ont eu un impact significatif dans le domaine.
— Article de référence : "Greedy Layer-Wise Training of Deep Networks" (2006)
— Article de référence : "Deep Sparse Rectifier Neural Networks" (2011)

Ceux là qui sont la référence dans l'extraction des descripteurs dans des graphes sont :

1. Jiawei Han : Jiawei Han est un chercheur renommé dans le domaine de l'extraction de connaissances à partir de données. Ses travaux ont porté sur l'extraction de motifs fréquents dans les graphes et l'utilisation de techniques d'apprentissage automatique pour extraire des descripteurs à partir de données graphiques.
 - Article de référence : "Mining Graphs, Time Series, and Streams" (2012)
 - Article de référence : "Graph-based social media analysis" (2009)
2. Christos Faloutsos : Christos Faloutsos est un chercheur spécialisé dans l'exploration de données complexes, y compris les graphes. Ses travaux ont porté sur la découverte de lois et de motifs structurels dans les graphes, ainsi que sur l'extraction de descripteurs pour la caractérisation et l'analyse des graphes.
 - Article de référence : "Graph Mining : Laws, Generators, and Algorithms" (2007)
 - Article de référence : "Power-law distributions in empirical data" (1999)
3. Lise Getoor : Lise Getoor est une chercheuse reconnue dans le domaine de l'apprentissage automatique et de l'extraction de connaissances à partir de données. Ses travaux se sont concentrés sur l'extraction de descripteurs à partir de graphes probabilistes et d'autres modèles de graphe.
 - Article de référence : "Link Mining : A Survey" (2005)
 - Article de référence : "Link-based classification" (2004)
4. Pedro Domingos : Pedro Domingos est un chercheur qui a contribué à plusieurs aspects de l'apprentissage automatique, y compris l'extraction de descripteurs dans les graphes. Ses travaux ont porté sur l'utilisation de modèles de graphe probabilistes et d'algorithmes d'apprentissage automatique pour extraire des informations significatives des données graphiques.
 - Article de référence : "Markov Logic : A Unifying Framework for Statistical Relational Learning" (2006)
 - Article de référence : "The Role of Occam's Razor in Knowledge Discovery" (1997)
5. Michalis Vazirgiannis : Michalis Vazirgiannis est un chercheur qui s'est intéressé à l'extraction de descripteurs dans les graphes et à l'analyse de données complexes. Ses travaux ont porté sur l'utilisation de méthodes de clustering et de classification pour extraire des descripteurs à partir de graphes, ainsi que sur l'exploration de motifs dans les données graphiques.
 - Article de référence : "Graph-based data mining : a survey" (2005)
 - Article de référence : "Clustering and community detection in directed networks : A survey" (2010)

9 Les conférences et journaux de référence des dieux dans le ML

Conférences :

1. Conference on Neural Information Processing Systems (NeurIPS) : NeurIPS est l'une des conférences les plus importantes dans le domaine de l'apprentissage automatique, mettant l'accent sur les aspects théoriques et pratiques de l'apprentissage automatique et des réseaux neuronaux.
2. ICML! (ICML!) : ICML est une conférence de premier plan dans le domaine de l'apprentissage automatique, couvrant un large éventail de sujets liés à l'apprentissage automatique, y compris les méthodes supervisées et non supervisées, l'apprentissage en ligne, etc.
3. Association for the Advancement of Artificial Intelligence (AAAI) Conference : AAAI est une conférence majeure dans le domaine de l'intelligence artificielle, qui comprend également des contributions importantes dans le domaine du machine learning.

4. International Conference on Learning Representations (ICLR) : ICLR est une conférence spécialisée dans la représentation de l'apprentissage automatique, en mettant l'accent sur les architectures de réseau profond, les embeddings et les méthodes de représentation avancées.

Journaux :

1. Journal of Machine Learning Research (JMLR) : JMLR est un journal en accès libre qui publie des articles de recherche de haute qualité couvrant tous les aspects du machine learning.
2. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) : TPAMI est un journal de l'IEEE qui se concentre sur les aspects théoriques et pratiques de la vision par ordinateur, du traitement d'images et du machine learning.
3. Journal of Artificial Intelligence Research (JAIR) : JAIR est un journal qui publie des articles de recherche approfondis dans le domaine de l'intelligence artificielle, y compris des contributions importantes dans le domaine du machine learning.
4. Machine Learning Journal (MLJ) : MLJ est un journal qui couvre un large éventail de sujets liés à l'apprentissage automatique, y compris les méthodes, les applications et les développements récents.

10 Axes de recherche dans le ML

Les axes de recherche se définissent comment les différentes problématiques auxquelles les gens s'intéressent. Dans notre domaine nous avons :

1. **Modèles d'apprentissage automatique** : Les chercheurs explorent de nouveaux modèles d'apprentissage automatique, tels que les réseaux de neurones profonds, les machines à vecteurs de support (Support Vector Machine (SVM)), les arbres de décision, les processus gaussiens, etc. Ils cherchent à développer de nouveaux modèles plus puissants, plus flexibles et plus adaptables pour résoudre une variété de problèmes.
2. **Interprétabilité et explicabilité** : L'interprétation et l'explication des modèles d'apprentissage automatique sont des thèmes de recherche importants. Les chercheurs travaillent sur des techniques permettant de comprendre comment les modèles prennent des décisions, d'expliquer leurs prédictions et de rendre les résultats plus transparents et interprétables pour les utilisateurs.
3. **Apprentissage en ligne et adaptatif** : L'apprentissage en ligne est un domaine de recherche qui se concentre sur la mise à jour continue des modèles à mesure que de nouvelles données deviennent disponibles. Les chercheurs explorent des techniques d'apprentissage en ligne efficaces et adaptatives pour gérer les flux de données en temps réel et s'adapter aux changements de distribution des données.
4. **Apprentissage non supervisé** : L'apprentissage non supervisé vise à découvrir des structures et des modèles cachés dans les données sans étiquettes. Les chercheurs explorent des algorithmes d'apprentissage non supervisé tels que la clustering, la détection d'anomalies, la réduction de dimension, etc., pour extraire des informations utiles et des représentations significatives des données.
5. **Apprentissage fédéré et confidentialité** : L'apprentissage fédéré est une approche qui permet d'entraîner des modèles sans partager les données brutes. Les chercheurs développent des techniques d'apprentissage fédéré qui préservent la confidentialité des données tout en permettant une collaboration efficace entre plusieurs parties.
6. **Apprentissage par renforcement** : L'apprentissage par renforcement est une branche du machine learning qui se concentre sur l'apprentissage à partir de l'interaction avec un environnement. Les chercheurs explorent des algorithmes d'apprentissage par renforcement pour améliorer la prise de décision séquentielle et l'apprentissage d'agents autonomes.

7. **Apprentissage en profondeur et vision par ordinateur** : L'apprentissage en profondeur (deep learning (DL)) est un sous-domaine du machine learning qui se concentre sur les réseaux de neurones profonds et leur application à des problèmes complexes tels que la vision par ordinateur, la reconnaissance vocale, la traduction automatique, etc. Les chercheurs explorent des architectures de réseau plus avancées, des techniques de pré-entraînement et des approches de transfert d'apprentissage pour améliorer les performances des modèles.

11 Tâches pour arriver au choix de l'axe de recherche

1. **État de l'art** : Il est important de se familiariser avec l'état actuel de la recherche dans le domaine du machine learning. Cela implique la lecture d'articles de recherche, la participation à des conférences et la prise de connaissance des travaux récents et des tendances émergentes.
2. **Identification des problèmes ou des lacunes** : Une fois que vous avez une bonne compréhension de l'état de l'art, identifiez les problèmes non résolus ou les lacunes dans le domaine qui vous intéressent particulièrement. Cela peut inclure des problèmes liés à l'efficacité, à la précision, à la scalabilité, à l'interprétabilité, à la confidentialité des données, etc.
3. **Domaines d'application** : Considérez les domaines d'application qui vous intéressent. Le machine learning est utilisé dans de nombreux domaines tels que la santé, la finance, l'ingénierie, la robotique, la vision par ordinateur, la reconnaissance de la parole, etc. Identifiez les défis spécifiques et les opportunités de recherche dans ces domaines.
4. **Collaboration et ressources** : Pensez aux opportunités de collaboration avec d'autres chercheurs, des institutions académiques ou des entreprises. Certaines pistes de recherche peuvent nécessiter des ressources spécifiques, des bases de données, des infrastructures informatiques ou des compétences particulières. Il est important d'évaluer les ressources disponibles pour soutenir votre recherche.
5. **Impact et pertinence** : Considérez l'impact potentiel de votre recherche. Pensez à la pertinence de votre travail pour la société, l'industrie ou la recherche fondamentale. L'identification d'un axe de recherche qui peut avoir un impact significatif peut être motivant et stimulant.
6. **Objectifs personnels** : Évaluez vos propres intérêts, compétences et objectifs de carrière. Choisissez un axe de recherche qui vous passionne personnellement et qui correspond à vos compétences et à vos aspirations professionnelles.
7. **Discussion et réflexion** : Discutez de vos idées avec vos pairs, vos mentors ou vos collègues. Participez à des séminaires, des ateliers ou des groupes de recherche pour partager vos idées et obtenir des commentaires. La réflexion et les échanges avec d'autres chercheurs peuvent vous aider à affiner votre axe de recherche.

12 Axe de recherche du sujet de recherche

13 Conclusion

Annexe

```
52
53 data_t data;
54 pthread_mutex_t mutex_sum; // déclaration du label de variable du
55                               mutex
56 void *produit_scalaire_parallele_1(void *arg)
57 {
58     /* Define and use local variables for convenience */
59     int i, taille;
60     long debut, fin;
61     long indice;
62     float som_locale, *x, *y;
63     indice = (long)arg;
64
65     taille = data.length;
66
67     debut = indice*(taille/NM_THREADS);
68
69     if(indice == (NM_THREADS - 1))
70         fin = data.length;
71     else
72         fin = debut + (taille/NM_THREADS);
73
74     x = data.x;
75     y = data.y;
76
77     /* Perform the dot product and assign result to the appropriate variable in the structure. */
78     som_locale = 0;
79     for (i=debut; i<fin; i++)
80     {
81         som_locale += (x[i] * y[i]);
82     }
83     //printf("Thread %ld debut = %ld fin = %ld\n", indice, debut, fin);
84
85     /* Lock a mutex prior to updating the value in the shared structure, and unlock it upon updating. */
86     pthread_mutex_lock (&mutex_sum); // verrouiller un mutex pour édition,
87     data.sum += som_locale;           puis éditer et enfin déverrouiller
88     pthread_mutex_unlock (&mutex_sum);
89
90     pthread_exit((void*) 0); // libérer le thread
91 }
92
93
```

FIGURE 1 – Verrou transactionnel entre threads

Références

- [Bis06] Christopher M Bishop, *Pattern recognition and machine learning*, Springer (2006).
- [GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, Nature **521** (2016), no. 7553, 436–444.
- [HS06] Geoffrey E Hinton and Ruslan R Salakhutdinov, *Reducing the dimensionality of data with neural networks*, Science **313** (2006), no. 5786, 504–507.
- [HZRS16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, *Deep residual learning for image recognition*, Proceedings of the IEEE conference on computer vision and pattern recognition (2016), 770–778.
- [KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, *Imagenet classification with deep convolutional neural networks*, Advances in neural information processing systems, 2012, pp. 1097–1105.
- [LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, *Deep learning*, Nature **521** (2015), no. 7553, 436–444.
- [PY09] Sinno Jialin Pan and Qiang Yang, *A survey on transfer learning*, IEEE Transactions on knowledge and data engineering, vol. 22, IEEE, 2009, pp. 1345–1359.
- [SB98] Richard S Sutton and Andrew G Barto, *Introduction to reinforcement learning*, MIT press (1998).
- [Vap95] Vladimir Vapnik, *The nature of statistical learning theory*, Springer Science & Business Media (1995).
- [VSP⁺17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, *Attention is all you need*, Advances in neural information processing systems (2017), 5998–6008.