

Deep Learning for Finance : Sujets de Projet et Consignes pour Master 2 EIF

Rémi Genet

April 3, 2025

Abstract

Ce document vous présente les sujets de projet pour le Master 2 en Deep Learning for Finance. Vous devez réaliser vos projets en groupes de quatre et respecter le format standard d'un article académique comprenant : Introduction, Motivation, Revue de littérature, Modèle, Expériences, Résultats et Conclusion. Le code devra être hébergé dans un dépôt GitHub public et les expériences exécutables via un Notebook Jupyter pour permettre une vérification directe des résultats.

1 Allocation des Groupes et Choix des Sujets

Vous devez effectuer vos projets en groupes de quatre. Chaque groupe devra choisir l'un des sujets définis ci-dessous :

- 1. Sig-TKAN pour la Prédiction du Volume ou de la Volatilité :**
Inspiré par le papier SigGate [1] (<https://arxiv.org/abs/2502.09318>), qui améliore les architectures GRU et LSTM en intégrant des signatures de chemin dans le mécanisme de porte, ce projet vise à appliquer la même idée dans le cadre du TKAN [2] (<https://arxiv.org/abs/2405.07344>). Je vous demande d'implémenter la boucle manuellement (c'est-à-dire sans utiliser la sous-classe RNN de Keras). Vous pourrez reprendre les mêmes tâches que celles proposées dans les articles, à savoir la prédiction du volume ou de la volatilité, ou bien choisir vous-même une autre tâche ou modifier le dataset. L'intérêt est de tester si l'intégration des signatures améliore la performance prédictive dans le TKAN également.
- 2. KAN Récurrent à Porte (Gated Recurrent KAN) :**
Basé sur le cadre TKAN [2] (<https://arxiv.org/abs/2405.07344>), qui intègre KAN avec un mécanisme récurrent de type LSTM, ce projet vous propose d'explorer une intégration similaire en utilisant une approche basée sur GRU. Vous pouvez reprendre les mêmes tâches que celles des articles (prédiction du volume ou de la volatilité), ou bien choisir vous-même une autre tâche ou modifier le dataset. Vous êtes invités à expérimenter

diverses méthodes d'intégration de KAN dans le mécanisme de GRU, et éventuellement modifier légèrement ce mécanisme. L'intérêt de cette approche réside dans l'exploration de la robustesse et de la flexibilité des architectures récurrentes adaptées aux problématiques financières.

3. Encodage pour l'Explication des Rendements :

Dans ce projet, vous utiliserez les cours de clôture ajustés quotidiens des actions du S&P 500 (de 2000 à 2025) afin de construire une architecture encodeur-décodeur. L'encodeur sera entraîné avec les données jusqu'en 2015 et le décodeur se limitera à une transformation linéaire simple. Pour la période 2015-2025, les représentations encodées seront utilisées dans une analyse de régression linéaire pour expliquer les rendements des actifs. Vous comparerez les performances de votre modèle aux facteurs de Fama et French [3] (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html) ainsi qu'aux résultats obtenus par une transformation PCA. Différentes architectures d'encodeur (linéaire, GRN, GRKAN ou toute autre approche non linéaire) devront être testées.

4. Implementation Shortfall :

Ce projet se focalise sur le modèle d'impact AFS pour un actif unique ([4] mais la formule est expliquée plus simplement dans [5] (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5046242), section 2). L'objectif est de développer une architecture RNN intégrant cette fonction d'impact afin d'obtenir un prix d'exécution optimal avec une déviation minimale par rapport au prix initial, en prenant en compte l'impact de marché. Vous pouvez vous inspirer des travaux sur le VWAP dynamique [6] (<https://www.arxiv.org/abs/2502.18177>). Pour simplifier, une période d'exécution maximale pourra être fixée et vous comparerez la performance du modèle à celle obtenue avec une exécution en TWAP sur cette période maximale, ainsi qu'à celle obtenue avec une exécution complète dès le premier pas de temps.

5. Estimation des Bid-Ask Spreads via Deep Learning :

Inspiré par l'article *Estimation of bid-ask spreads in the presence of serial dependence* de Xavier Brouty, Matthieu Garcin et Hugo Rocco (<https://arxiv.org/abs/2407.17401>), ce projet vise à estimer le spread bid-ask moyen quotidien à partir des données transactionnelles. Alors que l'article introduit des hypothèses de dépendance sérielle pour améliorer les estimateurs classiques, le projet se concentrera uniquement sur la calibration d'une fonction généralisable sur les données réelles, sans formuler d'hypothèses préalables sur la dépendance sérielle. Les entrées du modèle seront des barres d'une minute issues de Binance (<https://data.binance.vision/?prefix=data/futures/um/monthly/klines/>), et l'objectif sera de développer un modèle de deep learning capable, à partir des données d'un actif, de fournir une estimation fiable de son spread moyen de la journée - à calculer sur ces vrais valeurs: ([2](https://data.</p></div><div data-bbox=)

`binance.vision/?prefix=data/futures/um/monthly/bookTicker/`). Le modèle devra être entraîné sur plusieurs actifs simultanément sur une période donnée, puis testé sur une période ultérieure et sur un groupe d'actifs différent. Vous êtes invités à expérimenter divers modèles, incluant au minimum un MLP de base, ainsi que des approches utilisant des convolutions 1D et des RNN (LSTM, GRU, TKAN, etc.). Il est également conseillé d'explorer l'intégration du tick (la précision décimale des prix) en tant que caractéristique supplémentaire.

Note : Si deux groupes choisissent le même sujet et ne parviennent pas à se mettre d'accord, je vous demande de me contacter. Un tirage au sort sera réalisé pour trancher.

2 Structure du Document et Exigences

Votre projet devra être documenté dans un article rédigé en \LaTeX et respecter la structure suivante :

- **Introduction :** Présentez brièvement le problème de recherche et les objectifs du projet.
- **Motivation :** Expliquez l'intérêt de l'approche choisie dans le contexte du deep learning appliqué à la finance.
- **Revue de Littérature :** Résumez les travaux antérieurs et la littérature pertinente.
- **Modèle :** Décrivez en détail l'architecture proposée, en précisant les modifications ou intégrations novatrices (par exemple, l'utilisation des signatures ou l'intégration avec GRU). La description doit inclure les équations permettant de recréer le modèle.
- **Expériences :** Détaillez les jeux de données utilisés, le protocole expérimental et les méthodologies mises en œuvre.
- **Résultats :** Présentez et analysez les résultats expérimentaux. Il est important de noter que des résultats négatifs sont acceptables tant que la démarche scientifique est rigoureuse.
- **Conclusion :** Résumez les conclusions tirées et discutez des pistes d'amélioration ou des perspectives futures.

En complément, je vous demande de vous assurer que :

- L'ensemble du code est hébergé dans un dépôt GitHub public.
- Un Notebook Jupyter contenant les expériences est fourni, permettant une vérification directe des résultats.

3 Informations Complémentaires

- Le projet **Implementation Shortfall** est plus complexe et sera noté en conséquence.
- Une bonne recherche ne se mesure pas uniquement par des résultats positifs ; même des résultats négatifs constituent une contribution valide si la méthodologie et l'analyse sont bien réalisées.
- Si votre groupe souhaite explorer une idée de projet non listée ici, contactez-moi en détaillant votre idée et votre plan de travail afin de vérifier sa faisabilité.
- Si votre projet obtient des résultats prometteurs, n'hésitez pas à le soumettre sur SSRN (une démarche souvent plus simple que sur arXiv nécessitant une recommandation).

Références

References

- [1] SigGate Paper. Disponible à : <https://arxiv.org/abs/2502.09318>.
- [2] TKAN Paper. Disponible à : <https://arxiv.org/abs/2405.07344>.
- [3] Fama and French Data Library. Disponible à : https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.
- [4] AFS Impact Model Paper. Disponible à : <https://arxiv.org/pdf/0708.1756>.
- [5] Concave Cross Impact that explain easily AFS in section 2. Disponible à : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5046242.
- [6] Recurrent Neural Networks for Dynamic VWAP. Disponible à : <https://www.arxiv.org/abs/2502.18177>.
- [7] Brouty, X., Garcin, M., & Roccaro, H. *Estimation of bid-ask spreads in the presence of serial dependence*. Disponible à : <https://arxiv.org/abs/2407.17401>.