



Mars 2018
Wei GE
Yi QIAO

Optimisation d'un portefeuille d'actifs par les réseaux de neurones

Encadrant :

Jean-Marc LE CAILLEC
Didier GUERLOT



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Table des matières

1	Introduction	3
1.1	Problématique	3
1.2	Présentation de données	3
2	Conception et développement de modèle	4
2.1	Construction de la base d'apprentissage	4
2.2	Choix du modèle	4
2.3	Préparation des scénarios	4
3	Résultats obtenus et analyse	5
3.1	Résultats de différents scénarios	5
3.1.1	Variation de taille d'apprentissage	5
3.1.2	Variation de l'horizon	6
3.2	Evaluation du modèle	6
3.3	Difficultés rencontrées	6
4	Prospectives et points à améliorer	7
5	Conclusion	8

1 Introduction

Dans le marché financier, tout le monde cherche à gérer et optimiser son portefeuille pour obtenir le rendement attendu. Il y a une trentaine d'années, beaucoup de monde a commencé à essayer de prédire le rendement du marché de l'action en utilisant le volume et les prix historiques, le prix le plus haut, le plus bas, d'ouverture et de clôture. Plus tard, des indicateurs techniques (TIs), qui représentent la tendance, la volatilité et l'ampleur du marché, ont été développés et calculés à partir de volume et de prix pour rendre la prédiction plus efficace.

Dans le cadre de notre projet, nous travaillons sur les données historiques de TI des actions dans l'indice CAC40 pour prédire le rendement en appliquant un modèle de réseaux de neurones. Il y a trois facteurs principaux qui vont impacter la prédiction du rendement, l'horizon, la longueur des données historiques utilisées et le nombre de TIs.

Dans la vraie vie, les investisseurs souvent font des opérations (par exemple, achat ou vente d'une action) pour manipuler leur portefeuille et laissent passer le temps sans prendre aucune mesure. Alors que cela peut être un bon choix lors de l'achat, et peut devenir de moins en moins performant au fil du temps. Par conséquent, il est très important de décider l'horizon approprié entre les deux opérations en prenant en compte le risque de la volatilité du marché, sachant qu'il est coûteux de faire des opérations fréquemment.

Comme nous travaillons sur des séries temporelles, qui d'ailleurs ne sont pas stationnaires de longue mémoire, il est intéressant de pouvoir déterminer la durée de données d'apprentissage pour mieux prédire le rendement à un horizon prédéfini, ainsi que le nombre de TI.

Ce rapport est pour vous présenter notre projet en détail, qui contient 5 parties en global. D'abord, nous expliquons le contexte du projet, nous montrons notre problématique et les données fournies par nos encadrants. Ensuite, nous présentons le modèle de réseaux neurones pour réaliser la prédiction, qui consiste en la construction de la base d'apprentissage et les paramètres du modèle. Nous analysons les résultats obtenus et aussi évaluons notre modèle par les indicateurs statistiques. Dans la partie suivante, nous donnons des points à améliorer pour la continuation de ce projet. Ce rapport se termine par une conclusion synthétique.

1.1 Problématique

Un portefeuille d'actifs est construit par plein d'actions des poids différents, les actions sont impactées par plusieurs facteurs comme le secteur, le pays, etc. Pour gérer un portefeuille d'actifs, l'objectif est de maximiser le rendement global, c'est-à-dire qu'il faut être capable de faire la décision d'acheter ou de vendre des actions pour optimiser le rendement de chaque action. Dans le cadre de notre projet, nous avons besoin d'optimiser d'un portefeuille CAC40. Comme les séries temporelles sont non stationnaires, donc il est très difficile de faire la prédiction pour chaque période.

1.2 Présentation de données

Le portefeuille CAC40 est construit par 49 actions françaises et chaque action a ses comportements complètement différents. Nous avons un fichier qui contient le prix (ouverture, clôture, haut, bas) et le volume de chaque action pendant presque 15 ans (2000-2015). Les données dans ce fichier seront utilisées pour calculer le rendement d'un horizon déterminé. Pour une série temporelle, si nous prenons une période trop courte, nous n'aurons pas assez d'informations pour faire la prédiction. En revanche, elle n'a pas une forte corrélation pendant une longue période, les données très anciennes ne sont pas utiles de prédire la situation actuelle.

Nous avons aussi un fichier ayant les 13 « technical indicators » calculés à partir de ces séries temporelles. Les TIs se situent dans 4 catégories : le volume, la volatilité, la tendance et le momentum.

2 Conception et développement de modèle

Nous avons trouvé qu'il y a déjà beaucoup de projets sur la prédiction d'une séries temporelle financière. Pour atteindre notre objectif, nous avons proposé le modèle de réseaux neurones pour faire la prédiction. L'entrée et la sortie de notre projet sont très différentes par rapport aux autres projets. Les entrées sont les 13 TIs, comme nous savons plus de TIs peuvent apporter plus d'informations, mais il y a aussi plus de risks d'avoir trop de redondance. La sortie est le rendement sur un horizon prédéfini, c'est plus difficile de prédire la vraie valeur que la tendance (mettre un seuil d'achat ou de vente).

2.1 Construction de la base d'apprentissage

Pré-traitement de données :

Comme nous avons pris un apprentissage supervisé, nous avons besoin d'abord d'ajouter les labels pour nos exemples, c'est-à-dire nous avons calculé le rendement sur un horizon pour chaque exemple.

Dans un premier temps, nous avons utilisé directement les données non normalisées, parce que nous avons considéré que notre modèle peut ajuster les valeurs par lui-même. Cependant, quand nous avons fait le premier test sur les données non normalisées, nous avons trouvé que le résultat n'était pas idéal. Le système n'a donné que un constant négatif et un constant positif comme le résultat, il n'a pas fait une l'approchement avec la vraie valeur. Ensuite, nous avons fait une normalisation sur les entrées et nous avons pris le formulaire $(V - V_{min}) / (V_{max} - V_{min})$, c'est-à-dire que nous avons mis toutes les valeurs de l'entrée sur la fourchette de $[0,1]$. Après la normalisation, nous pouvons trouver que le résultat était mieux.

Paramétrage de la base de données :

Pour notre projet, il faut aussi varier sur les paramètres différents pour construire la base d'apprentissage. Principalement, nous avons pris 4 indicateurs pour mesurer notre système, ils sont l'horizon, la taille d'apprentissage, la durée du test et le nombre de TI.

Pour lancer les simulations différentes, nous avons d'abord changé seulement un indicateur sur les tests. Et puis, nous avons testé sur le cas où nous pouvons changer 2 ou 3 indicateurs.

Nous avons pris tous les 13 TI au début de notre projet, nous voudrions combiner plus d'information pour faire la prédiction, nous avons considéré que cette combinaison de TI est plus robuste. Ensuite, nous avons diminué le nombre de TI pour savoir le performance de notre système. Nous avons refait les tests avec la dimension réduite de TI.

2.2 Choix du modèle

2.3 Préparation des scénarios

3 Résultats obtenus et analyse

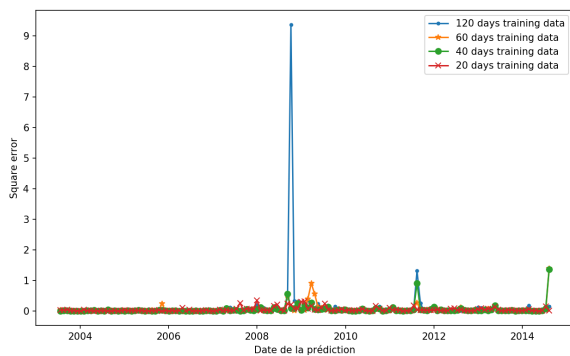
3.1 Résultats de différents scénarios

3.1.1 Variation de taille d'apprentissage

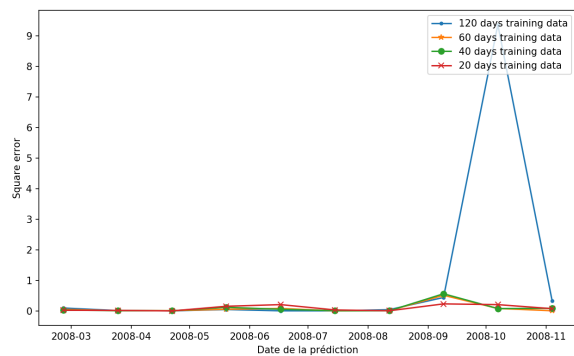
Comme nous travaillons sur des séries temporelles, qui ne sont pas stationnaires, nous nous intéressons à l'effet de la taille d'apprentissage sur la précision de la prédiction du rendement de l'action. Nous avons réalisé des tests sur l'action Renault dans l'indice CAC40.

Sur la figure 1, nous observons un très grand pic au début d'octobre 2008 avec une taille d'apprentissage de 120 jours (6 mois), juste après le début de la crise mi-septembre. Toutefois, ce pic est absent sur les autres courbes. Cela signifie que notre modèle est capable de détecter le début de la période de la crise, puisqu'une fois que celle-ci est arrivée, la tendance du marché est brutalement rompue. Dans ce cas, apprendre sur les données des 6 derniers mois n'est pas une bonne approche pour la prédiction du rendement. En revanche, apprendre sur une période plus courte est plus efficace pour ne prendre en compte que les informations intéressantes de l'évolution du marché.

Dans les autres périodes, nous pouvons constater que l'erreur quadratique de la prédiction du rendement est proche de zéro dans les quatre cas.



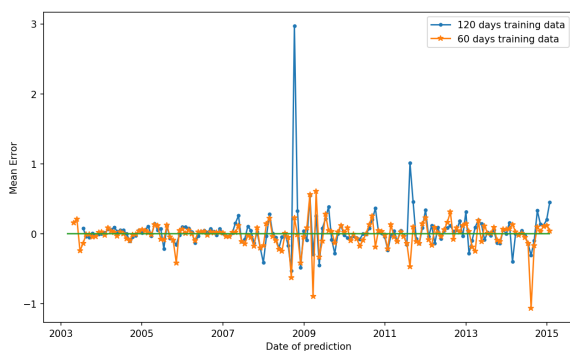
(a) Vue globale sur 10 ans



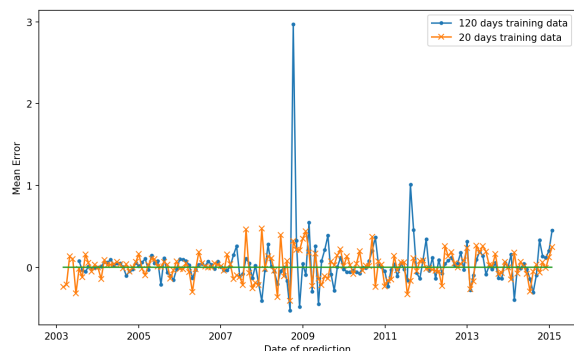
(b) Vue zoomée sur la période de la crise de 2008

FIGURE 1: Erreur quadratique de la prédiction du rendement de Renault à un horizon de 20 jours (1 mois) en fonction de la taille d'apprentissage

En observant les figures ci-dessous, nous remarquons que l'erreur moyenne pour 120 jours est généralement plus petite que les autres cas (sauf la période de la crise), car une période plus longue contient plus d'informations et peut mieux présenter la tendance du marché. Au contraire, une période plus courte est plus bruitée.



(a) Erreur moyenne pour 120 et 60 jours



(b) Erreur moyenne pour 120 et 20 jours

FIGURE 2: Erreur Moyenne de la prédiction du rendement de Renault à un horizon de 20 jours (1 mois) en fonction de la taille d'apprentissage

3.1.2 Variation de l'horizon

3.2 Evaluation du modèle

3.3 Difficultés rencontrées

L'apprentissage a pris trop de temps.

Trop de scénarios

Bcp d'actions

4 Prospectives et points à améliorer

La décision (à seuil, achat, vente, ne rien à faire)
Appliquer à toutes les actions dans le portefeuille

5 Conclusion