

Gestion de portefeuille avec les algorithmes d'apprentissage

Responsable du projet :

Yoshua Bengio

Département d'informatique et recherche opérationnelle et Centre de Recherches Mathématiques

Université de Montréal

Tél. : (514)343-6804

Fax : (514)343-5834

Courriel : bengioy@iro.umontreal.ca

Chercheurs principaux :

René Garcia

Département de sciences économiques et CIRANO

Université de Montréal

Tél. : (514)343-5960

Fax : (514)343-5831

Courriel : garciar@cirano.umontreal.ca

Description du projet

Motivation

En ce qui concerne les travaux des deux prochaines années, nous voulons en particulier étudier la question de la confiance qu'on peut accorder aux décisions d'un algorithme d'apprentissage. Ceci peut aider le gestionnaire de portefeuille à faire un choix rationnel entre les différentes méthodes, ou bien à tenir compte de l'incertitude sur les décisions d'un algorithme pour mieux contrôler ses risques. Comment peut-on comparer différentes méthodes ? On peut faire une simulation sur les données historiques et mesurer la qualité des décisions qui auraient été prises, hors-échantillon, si on avait suivi chacune des méthodes. Cela nous donne un estimateur non-biaisé mais bruité de la qualité de ces méthodes (par exemple, il ne suffit pas de mesurer la performance d'un gestionnaire à partir de quelques décisions pour évaluer de manière fiable la qualité de ses décisions). Pour savoir si une nouvelle méthode vaut vraiment la peine d'être utilisée, il faut aussi avoir une idée de l'incertitude qui existe dans notre estimation de sa performance. De tels intervalles de confiance seraient aussi utiles pour savoir si on devrait suivre une recommandation particulière faite par une méthode à un moment donné plutôt que de prendre une décision conservatrice (tel que ne pas faire de transaction ou prendre une position moins risquée).

Algorithmes d'apprentissage et intervalles de confiance

Les algorithmes d'apprentissage sont des méthodes permettant d'exploiter des relations

entre des variables dont on ne connaît pas la distribution jointe mais pour lesquels on observe un nombre fini d'exemples (ce sont les données, ici des séries financières et économiques). Nous travaillons en particulier sur les réseaux de neurones artificiels, qui permettent d'apprendre des relations non-linéaires entre des variables (entre des variables d'entrée et des variables de sortie), sans faire de suppositions explicites sur la forme de la distribution de ces variables. Dans les deux dernières années nous avons développé plusieurs algorithmes d'apprentissage pour prendre des décisions financières ou faire des prédictions, et certaines de ces méthodes sont utilisées par nos partenaires privés avec succès. Maintenant, nos partenaires nous demandent de leur fournir des outils permettant d'évaluer de la manière la plus juste possible la confiance qu'on peut accorder aux décisions proposées par ces méthodes, et la confiance avec laquelle on peut affirmer qu'une méthode est généralement plus utile qu'une autre. Dans le contexte d'une gestion rationnelle des risques, ces questions sont effectivement très importantes. D'un point de vue théorique, il apparaît aussi que si on a plusieurs méthodes pour lesquelles on a des intervalles de confiance sur les prédictions, on peut améliorer les performances en combinant les prédictions ou les décisions de ces méthodes (en tenant compte de la confiance qu'on accorde à chacune). Dans le contexte des applications classiques des algorithmes d'apprentissage (où les données sont supposées *iid*, identiquement et indépendamment distribuées), on peut par exemple obtenir de tels intervalles avec le Bootstrap non-paramétrique. Malheureusement, à cause des dépendances temporelles et des non-stationarités, ces méthodes ne s'étendent pas facilement aux séries chronologiques. Une des méthodes préférées, le Bootstrap par blocs, ne fonctionne pas bien quand la longueur des dépendances est grande par rapport au nombre total de points dans la séquence historique, ce qui est le cas pour les données mensuelles.

Objectifs

Les objectifs de la recherche proposée sont donc doubles. Dans un premier temps il s'agit de mettre au point des méthodes permettant d'obtenir des estimations d'intervalles de confiance de certaines fonctions non-linéaires des données (tel que l'erreur mesurée hors échantillon pour un algorithme d'apprentissage appliqué à une série chronologique). Dans un second temps, il s'agit d'exploiter de telles méthodes pour fournir des intervalles de confiance, diminuer les risques pris dans les décisions de gestion de portefeuille, et même améliorer la qualité des décisions en combinant plusieurs méthodes pour lesquelles on a des intervalles de confiance.

Méthodologie

En ce qui concerne la mise au point d'intervalles de confiance, l'approche proposée est basée sur l'utilisation d'algorithmes d'apprentissage faisant de l'estimation de densité conditionnelle pour modéliser les variations plausibles que la série historique aurait pu subir, en tenant compte des dépendances temporelles, et de s'en servir pour mettre au point de nouveaux algorithmes de Bootstrap pouvant tenir compte de longues dépendances dans les séries chronologiques.

Nous nous proposons d'exploiter deux types d'algorithmes pour modéliser les dépendances temporelles afin de générer des historiques plausibles. Le premier type

de méthode nous donnera ce qu'on appelle un Bootstrap paramétrique, basé sur un modèle probabiliste paramétrique des dépendances temporelles: les IOHMMs ("input/output hidden Markov models"). Ces modèles sont similaires aux modèles de Markov cachés utilisés en reconnaissance de la parole et aux "Markov switching models" développés en économétrie pour tenir compte des changements de régime dans les séries économiques. Nous venons de conclure la recherche d'une subvention stratégique portant justement sur l'utilisation de ce type d'algorithme d'apprentissage pour modéliser les changements de régime dans les données financières, et nous pourrions réutiliser pour le projet proposé certains étudiants diplômés et assistants de recherche formés dans le cadre de cette subvention stratégique. Le modèle a plusieurs états, non-observés, et dans chaque état la relation entre les variables d'entrée et les variables de sortie est modélisée, par exemple avec un réseau de neurones (dont les sorties donnent les paramètres de la distribution conditionnelle de la variable de sortie étant donné la variable d'entrée).

Le deuxième type de méthode que nous proposons d'utiliser nous donnera ce qu'on appelle un Bootstrap non-paramétrique, mais sans utiliser l'hypothèse que les données sont *iid*. Notons $y_1 \dots y_t$ la séquence historique (avec y_t un vecteur réel). On utilisera une extension des modèles non-paramétriques à noyaux qui permet de représenter la distribution conditionnelle $P(y_t|x_t)$ de deux variables aléatoires y_t et x_t . Un modèle à noyau classique de la densité peut s'exprimer comme la convolution de la distribution empirique (les observations passées) et du noyau (par exemple une densité gaussienne). Cela donne en fait une mixture (par exemple de gaussiennes centrées sur les observations passées). Dans notre cas, le modèle servira à représenter la distribution **conditionnelle** de l'observation y_t au temps t étant donné une fonction $x_t = f(y_1, \dots, y_{t-1})$ des observations passées. Le modèle non-paramétrique représente cette distribution conditionnelle $P(y_t|x_t)$ comme une mixture sur toutes les observations passées, comme dans le cas inconditionnel, mais avec des poids qui dépendent du contexte: on a plus de poids sur les observations qui ont été observées dans un contexte x , qui ressemble au contexte présent x_t . Pour définir la notion de similarité entre deux contextes x_t et x , on utilise un deuxième noyau. Un des problèmes qui se posent avec les modèles non-paramétriques est le choix du niveau de lissage des noyaux (ce qui correspond à la variance dans le cas d'un noyau gaussien). Pour ce faire nous voulons explorer plusieurs méthodes, la plus simple étant de maximiser la vraisemblance hors-échantillon qui aurait été obtenue si on avait choisi une valeur particulière pour ces paramètres de lissage. Finalement, nous proposons d'utiliser des techniques empruntées aux "Support Vector Machines" et aux algorithmes de "Matching Pursuit" pour sélectionner un sous-ensemble des données passées qui permettent de mieux représenter la densité.

Une fois que nous avons appris un modèle probabiliste, nous pouvons l'utiliser pour générer des séquences de données similaires à la séquence historique. Pour chacune des séquences générées, on obtient par exemple une prédiction, une décision, ou une moyenne des performances hors-échantillon d'un système de décision. On peut ensuite utiliser la variance échantillonnale ou la distribution empirique de ces résultats pour obtenir des intervalles de confiance, soit sur les décisions, les prédictions, ou les erreurs de généralisation. Cette information peut être utilisée de trois manières: (1)

simplement pour permettre au gestionnaire de portefeuille de prendre une décision particulière plus éclairée (en connaissant l'incertitude sur la prévision du modèle, par exemple), (2) pour lui permettre de comparer deux méthodes (grâce aux intervalles de confiance sur la performance), et (3) pour construire de meilleurs systèmes de décision, obtenus par une combinaison linéaire des décisions que donnent plusieurs méthodes (avec des poids qui dépendent de la confiance relative qui est accordée à chacune de ces méthodes). Dans les trois cas, on utilise le Bootstrap pour quantifier les variations plausibles autour de la prédiction ou de l'erreur obtenue avec l'algorithme d'apprentissage entraîné sur les vraies données historiques. En effet il est possible que le modèle probabiliste génératif nous donne un estimateur biaisé de l'espérance qui nous intéresse, donc nous proposons d'utiliser seulement les écarts à la moyenne empirique sur les différentes simulations du Bootstrap pour construire nos intervalles de confiance autour de la décision ou de l'estimateur habituel d'erreur pour l'algorithme d'apprentissage.

Réseautage

Ce projet représente une collaboration CRM-CIRANO et a aussi des liens très clairs avec d'autres projets de recherche du rcm_2 ainsi qu'avec d'autres réseaux comme le réseau MITACS. En ce qui concerne le rcm_2 , ce projet est lié de près, au niveau des techniques utilisées (mais pas des applications), aux projets de finance quantitatives du CIRANO dirigés par René Garcia (Sciences Économiques). Les étudiants de Yoshua Bengio développent des outils informatiques génériques, pour effectuer des simulations avec les algorithmes d'apprentissage, qui servent aussi bien dans les applications en finance que dans les applications en data-mining. Ce projet est ainsi aussi lié aux projets MITACS dans le domaine des technologies de l'information et en particulier celui sur l'inférence en haute dimension, qui s'attaque aussi à des problèmes de grande taille qui se retrouvent dans les applications de data-mining, en utilisant aussi des modèles probabilistes de la distribution jointe d'un grand nombre de variables. Dans le cas des séries financières cet ensemble de variables est la séquence elle-même.

Rapport d'étape

Le monde de la finance subit ces dernières années une transformation profonde qui affecte grandement le travail quotidien des gens qui y œuvrent. En effet, il apparaît de plus en plus que les outils des mathématiques, des statistiques, et de l'informatique sont essentiels pour pouvoir gérer des décisions financières aussi bien ou mieux que les autres grands acteurs de la finance internationale. L'information financière et économique est de plus en plus facilement accessible, et il y en a de plus en plus. Ce qui fait la supériorité d'une entreprise, dans ce contexte, c'est la capacité de mieux traiter ces données pour en extraire les connaissances pertinentes pour prendre de meilleures décisions. En particulier, les gestionnaires de portefeuille doivent choisir où placer leur actif. Pour prendre cette décision, ils doivent non seulement choisir les placements qu'ils croient les plus profitables mais aussi simultanément contrôler les

risques encourus. Ce projet s'inscrit donc dans le thème de la gestion du risque. Il existe beaucoup d'algorithmes qui partent des données historiques et qui mènent à une fonction de décision. Durant les deux dernières années nous avons développé de nouveaux algorithmes d'apprentissage, en particulier dans le domaine des réseaux de neurones, que nous avons appliqués aux problèmes de gestion de portefeuille de nos partenaires.

Progrès réalisés

Dans la section Publications et autres résultats on trouvera une description des fonctionnalités qui ont été livrées dans le logiciel (AD) que nous développons pour nos partenaires privés. Au niveau scientifique, les résultats les plus notables sont les suivants (voir numéros des publications plus bas):

- Nous avons mis au point une méthode originale pour optimiser ce qu'on appelle des hyper-paramètres dans les algorithmes d'apprentissage [2,6,7], que nous avons appliquée avec succès au problème de la prédiction de la volatilité future en tenant compte des non-stationarités [8,9,11].
- Nous avons étudié les algorithmes de modélisation de la densité de séries financières avec des modèles Markoviens [3] que nous avons appliqués à des séries financières de rendements [12]. Nous étendons aussi ces idées à la modélisation de distributions jointes d'un grand nombre de variables [10].
- Nous avons mis au point deux nouvelles méthodes pour faire plus correctement de l'inférence statistique sur l'estimation de l'erreur de généralisation d'algorithmes d'apprentissage [5].

- Nous avons démontré que les tests d'hypothèse classiques pour les paramètres d'une régression étaient trop libéraux quand ce qui nous intéresse est l'erreur de généralisation hors-échantillon sur des séries financières, et nous avons proposé des tests alternatifs [13].

Étudiants et stagiaires encadrés 11/98-11/99

Charles Dugas	étudiant au PhD, dept. IRO.
François Gingras	étudiant au PhD, dept. IRO, maintenant gestionnaire de portefeuille à Hydro-Québec.
Nicolas Chapados	étudiant au PhD, dept. IRO.
Vincent-Philippe Lauzon	étudiant à la maîtrise, dept. IRO, maintenant chercheur en data-mining pour Curvu en Australie.
Réjean Ducharme	assistant de recherche, CIRANO.
Samy Bengio	assistant de recherche, CIRANO, maintenant directeur du groupe de recherche sur les algorithmes d'apprentissage à l'IDIAP en Suisse.
Claude Nadeau	stagiaire post-doctoral, CIRANO, maintenant chercheur pour Statistiques Canada.

Publications et autres résultats

Nous avons créé un outil de simulation d'algorithmes d'apprentissage pour la finance, appelé AD (pour Adaptive Deciders). Avec ce logiciel, nous mesurons la performance en généralisation des systèmes de décision en tenant compte de la nature séquentielle des données financières et en faisant "comme si" un agent imaginaire avait accès à chaque instant dans l'histoire uniquement aux informations passées pour prendre une décision concernant la prochaine période. Cela semble une approche évidente, mais les critères usuels d'analyse et d'évaluation utilisés en statistique classique ne mesurent pas cela. Cette approche est très importante car il est facile de trouver des règles de décisions qui fonctionnent bien sur l'historique mais qui ne généralisent pas bien sur de nouvelles données. Pour cela il suffit de choisir une telle règle parmi un ensemble suffisamment grand de règles. Malheureusement, beaucoup des systèmes quantitatifs proposés dans le commerce souffrent de ce problème, aussi appelé "data-snooping". C'est-à-dire que la performance d'un système de décision a été mesurée sur des données qui ont été utilisées pour choisir ce système de décision! Parfois, le biais optimiste ainsi introduit est petit, mais quand on prend des décisions financières, il est crucial d'avoir une mesure honnête de performance. Dans AD, l'utilisateur peut spécifier un ensemble de modèles et le logiciel évaluera la performance en faisant une simulation réaliste, c'est-à-dire qu'à chaque période, un modèle sera choisi pour la prochaine décision en comparant les modèles entre eux sur les données passées. Durant la dernière année, les fonctionnalités suivantes ont été rajoutées au logiciel AD: modules pour entraîner des modèles probabilistes (incluant des modèles de Markov,

des modèles ARCH et GARCH, et des filtres de Kalman), algorithmes de sélection de variables (sélection par étape pour la régression linéaire et la régression non-linéaire par réseaux de neurones), système permettant d'optimiser les poids d'un portefeuille de gestion de portefeuille pour maximiser un ratio profit/risque en utilisant les prédictions d'un réseau de neurones.

Au niveau de la recherche, ce projet a généré les publications suivantes:

1. Schwenk, H. et Bengio Y. (1999), *Boosting Neural Network*, à paraître dans Neural Computation.
2. Bengio, Y. (1999), *Gradient-Based Optimization of Hyper-Parameters*, à paraître dans Neural Computation.
3. Bengio Y. (1999), *Markovian Models for Sequential Data*, Neural Computing Surveys, **2**, pp. 129 162.
4. Bengio S., Bengio Y., Robert J. et Bélanger G. (1999), *Stochastic Learning of Strategic Equilibria for Auctions*, Neural Computation, **11**(5), pp.1199 1209.
5. Nadeau, C. et Bengio, Y. (1999), *Inference for the Generalization Error*, à paraître dans Advances in Neural Information Processing Systems **12**.

ainsi que les présentations suivantes à des conférences ou des ateliers:

6. Bengio, Y., Latendresse, S., and Dugas, C. (1999), *Gradient-Based Learning of Hyper-Parameters*, Learning Conference, Snowbird, Utah.
7. Bengio, Y. (1999), *Yet Another Way to Defy the Curse of Dimensionality: Gradient-Based Learning of Hyper-Parameters*, presented at the Large Data Sets workshop, May 6-8, Waterloo, Ontario.
8. Bengio, Y., (1999), *Estimating and Improving Generalization Error for Non-IID Data*, presented at the Fields Institute, Toronto, Ontario, 4 Mai 1999.
9. Garcia R. et Bengio, Y. (1999), *Portfolio Management with Artificial Neural Networks*, aux F,4ME Research Days 1999, Genève, Suisse, 16-20 Avril 1999.
10. Bengio- Y. (1999), *Learning from Structured High-Dimensional Data*, rencontre de la Mathematical Society of Canada, 11-13 décembre 1999, Montréal.
11. Garcia, R. (1999), *A Monte Carlo Method for Optimal Portfolios*, Mathematical Finance Conference, HEC-GERAD, Montréal.

Et les articles suivants ont été soumis pour publication dans des revues scientifiques (et des rapports techniques correspondants ont été écrits):

12. Bengio, Y. et Dugas, C. (1999), *Forecasting Non-Stationary Volatility with Hyper-Parameters*, soumis à Machine Learning.
13. Bengio, Y., Lauzon V-P., et Ducharme R. (1999), *Experiments on the Application of*

IOHMMs to Model Financial Returns Series, accepté à IEEE Transactions on Neural Networks.

14. Gingras, F., Bengio, Y. et Nadeau, C. (1999), *On Out-of-Sample Statistics for Financial Time-Series*, publié à Computational Finance 2000.

Autres documents de travail

15. Detemple, J., R. Garcia et M. Rindisbacher (1999), *A Monte Carlo Method for Optimal Portfolios*, manuscrit, CIRANO.