



Start Up
Projet d'entrepreneuriat 2013

QuanTrade
Paris/Londres/Worldwide

Introduction au projet

Projet QuanTrade - Quantitative and collective trading intelligence

Xavier Bruhiere





Résumé :

Le petit projet en C initial pèse maintenant plusieurs mégas de maths et de code et scannera dans quelques heures, pour la première fois, sur mon serveur dédié, le marché mondial à la recherche de signaux d'opportunités de trades. Si j'écris donc ce document c'est pour passer à la vitesse supérieure et amener d'autres collaborateurs sur ce délire ambitieux et passionnant ! Cher lecteur avisé, si tu es un sournois personnage qui veut violemment exploiter mon dur labeur, poursuis ta lecture je pousse la naïveté jusqu'à détailler tous mes plans.

Je vais enfoncer des portes ouvertes pour certains, sautez à la section qui vous botte.

Mais puisqu'il s'agit de finance quantitative, et pour faire bien, je case d'abord une citation de Max Dama, auteur d'un des sites de trading les plus reconnus :

"Quantitative trading is the job of programming computers to trade. It is the best job out of undergraduate - you commonly make 100k base salary, plus a bonus which is often larger, the hours are flexible as long as you produce, there is no dress code, the work is extremely fascinating, fast paced, and uses the most advanced technology, and your performance is evaluated objectively."

Quant trading combines programming and trading. Programming is the fastest way to take an idea and turn it into reality (compare it to writing a book, publishing, etc or to architecting a building, getting zoning permits, doing construction, etc). Trading is the most direct approach to making money (compare it with starting the next facebook, programming, hiring, advertising, etc or to creating a retail chain, hiring, waiting for money to flow in to reinvest, etc). Combining programming and trading, you have the most direct path from an idea in your brain to cash."

Look at Forbes 400 (Mar 2011). Traders are disproportionately well-represented (so are trust funders). (Note that this is only the US. In many countries such as Russia, corruption is a better career path.)"

Quant trading is intellectually very satisfying because it draws heavily from fields as diverse as computer science, statistics, math, psychology, economics, business, operations research, and history."

Cool ! Après l'accroche vient l'annonce du plan : contexte du projet, explication de l'application de trading, pour conclure avec la roadmap.

Pas de business plan ni d'étude de marché dans ces quelques pages, mais plutôt un snapshot de ma réflexion et de mon travail actuel, pour convaincre de rejoindre l'équipe.



Table des matières

1	Projet	3
1.1	Data	3
1.2	Formatage	3
1.3	Enrichissement	3
1.4	Application	4
1.5	Visualisation	4
1.6	Communauté	5
1.7	Self-amélioration	5
2	Application	6
	Conclusion	8



1 Projet

La réflexion donc. La partie 2 est plus fun mais pour mieux cerner le potentiel du projet, je vais détailler les fragments successifs à haute valeur ajoutée d'une application de traitement de donnée moderne. Ce domaine applicatif est un leitmotiv des développements informatiques actuels, centrés sur la data dont le flux s'accroît en permanence. Associé à l'augmentation des capacités de stockage et de calcul des ordinateurs et à leur accessibilité, il s'agit d'une formidable opportunité pour qui a les connaissances requises.

L'objectif dans les paragraphes suivants est donc d'appréhender le pipeline que suit une donnée et de juger des étapes sur lesquelles apporter une solution plus puissante et innovante.

1.1 Data

Logiquement, le premier facteur clé est de la récupérer. L'open-data est à la mode et les sources gratuites de qualité émergent. QuanDL, twitter, marketdata, TrueFX, google finance, New-York Times... proposent tous une API dédiée à l'accès temps réel et gratuit de leurs données. L'enjeu est donc d'être capable d'identifier les bonnes sources d'information et de les exploiter.

1.2 Formatage

La multiplication des sources a son revers : pertinence et fiabilité non-systématique, quantité indigeste, doublons, synchronisation... Un premier travail qui fait vivre nombre de sociétés de service est donc d'appliquer une première transformation sur ce flux en le filtrant, nettoyant et normalisant, puis en le stockant dans des infrastructures assurant disponibilité rapide et permanente de quantités conséquentes et hautement dynamiques. Cette étape permet déjà de mettre à la disposition de communautés scientifiques compétentes de la matière sur laquelle produire des résultats valorisant (monétairement, intellectuellement, voir plus (tips : le médical)). Une perspective qui prend tout son sens lorsque l'on sait que les scientifiques spécialisés déclarent consacrer moins de la moitié de leur temps sur l'analyse proprement dite, du fait de ce nettoyage chronophage qu'ils mènent eux-même, pour l'instant.

1.3 Enrichissement

"any fool can know... the point is to understand." - Albert Einstein

Classiquement (quoique souvent amputé de l'étape précédente) tout ça est tranquillement branché en direct sur les applications utilisatrices. C'est gacher une opportunité! Toute cette quantité d'information peut - doit - être passée à la moulinette des dernières techniques de data mining, d'intelligence artificielle, d'algorithmes évolutionnistes, ... pour extraire d'avantage que la signification brute. Pattern des cours de bourse en regardant les années précédentes, prédiction de comportement depuis tous les tweets, fiabilité des informations en comparant tous les blogs, aide à la décision dans des systèmes à milliers de variables.

Avec cet enrichissement supplémentaire, on se fait les partenaires de premier choix des



applicatifs de traitement de données (soit une immense majorité grandissante du secteur), des laboratoires de recherche et d'autres institutions.

1.4 Application

Arrive donc lesdites applications exploitant ce flux. Sans surprise la tendance est aux cluster de bêtes de courses accessibles dans le cloud par des clients légers. Et c'est pertinent : quoi de mieux que de simuler en 24s 10 années de stratégie algorithmique de trading depuis sa tablette qui habituellement peine à décoder youtube. Avec la mise à disposition des données, c'est le second atout à glisser dans les mains de thésards chercheurs ou utilisateurs qui ne demandent pas mieux pour mettre en oeuvre et rentabiliser leurs compétences.

Pour prendre un exemple au coeur du sujet, le site quantopian.com qui propose de simuler des stratégies de trading obtient un formidable retour des communautés financières et scientifiques, qui toutefois doivent, pour les meilleurs, construire eux mêmes depuis les simples cours des données plus complexes. Ayant travaillé avec eux je sais que les concepteurs sollicitent de nouvelles sources de data plus sophistiquées.

1.5 Visualisation

Si on a bien travaillé jusqu'ici, on obtient de beaux résultats tout neufs (hum au hasard, un portefeuille d'action qui performe à 43%). Mieux vaut donc les voir. L'enjeu est d'acquérir une véritable compréhension des problèmes complexes étudiés et des résultats, afin d'enrichir la réflexion et d'inférer efficacement sur la suite des événements (avoir réussi à analyser les liens sociaux de facebook et ne pas pouvoir se dépêtrer de l'immonde graphique qu'a pondue l'algorithme est sans doute frustrant).

Coup de chance, l'avènement de l'ère de la data a poussé de nombreux développeurs à créer des outils de visualisation qui explosent aujourd'hui avec l'arrivée du HTML5 et de l'omniprésence du javascript.

A cela s'ajoute l'esthétisme dont l'intérêt en terme de communication et de diffusion n'est plus franchement à démontrer, pas plus qu'en matière d'ergonomie.

Pour les sceptiques : <http://orbifold.net/html5/> et www.d3js.org



FIGURE 1 – portfolio visualization



1.6 Communauté

Je viens de parler de diffusion et j'ai évoqué deux fois la mise à disposition des puissances processeurs et des capacités de stockage. L'idée en effet est d'impliquer les communautés étudiantes, open-sources, passionnées. . . Pourquoi ? Parce que les meilleurs cours de la planète sont suivis gratuitement sur la base du volontariat par des millions d'étudiants, que l'émulsion intellectuelle sur Quantopian.com est méchamment productive, que les projets open-source permettent de faire partir ses codes sur la base des meilleurs talents informatiques, que tous les jours les meilleurs statisticiens partagent leurs connaissances sur r-bloggers.com. Et parce que quand on est patron on aime avoir une équipe internationale de milliers de développeurs qui ne demandent pas d'autre salaire qu'un projet passionnant.

1.7 Self-amélioration

Et voilà qui nous amène au dernier point du process. Le travail de ces communautés, efficace grace aux outils cités et à la collaboration, doit être réinjecté dans ces fameux outils , qui d'eux même s'amélioreront en bouclant l'analyse !

Ce qu'il faut retenir : Pendant que la guerre fait rage sur les applicatifs et services informatiques, allons nous plonger en amont en interceptant, enrichissant, et refourguant de la donnée évoluée, et en aval pour valoriser leurs résultats. Et enfin impliquons les communautés open-source pour diminuer les frais et augmenter l'intelligence et la créativité des solutions.

Et tant qu'on y est développons le tout sur un système de trading que j'ai déjà écrit qui servira de fer de lance communiquant et crédible, et produira les bonus de début, milieu et fin de mois.



2 Application

Tout ce qui a été raconté jusqu'ici je l'ai découvert et expérimenté au cours du développement de QuanTrade. On va donc passer à la pratique en détaillant le fonctionnement du programme dans son état actuel et établir ce qui constitue les bases (du lourd les bases) du projet.

Deux précisions avant d'avancer : tout ce qui va être évoqué est fonctionnel, et est autant valable en temps réel qu'en backtest sur les années précédentes.

Le principe du programme est de surveiller régulièrement l'état du marché et procéder à des calculs pour détecter les opportunités d'achat ou de vente, de passer ces signaux à un manager de portefeuille qui proposera les ordres à émettre, selon diverses optimisations. Ce fonctionnement de base est hautement configurable : les algorithmes, le marché, la monnaie, le type de données utilisées... tout.

A partir de là plusieurs interactions utilisateurs sont possibles.

- La première est de simplement choisir parmi les listes les paramètres que l'on souhaite et regarder le système faire le sale boulot à sa place depuis un client web temps réel spécialisé en visualisation de donnée (Pour ceux qui ont suivi on commence à faire allusion à l'étape 5 de la partie précédente).
- Plus puissant et plus geek, il est possible d'écrire soi même un certains nombres de modules qui s'intégreront automatiquement au système. Ainsi le programme met à dispositions de puissants outils pour personnaliser les stratégies algorithmiques, les gestions de portefeuille, les optimisateurs de paramètres, les sources de données utilisées et bien sûr la façon de les combiner. L'ensemble peut donc également être vu comme un véritable labo de trading quantitatif (toujours pour ceux qui suivent l'histoire des sources de données est le point d'entrée des étapes 1, 2 et 3).
- Enfin la place de contributeur amène à intervenir sur l'ensemble du code et de la stratégie de développement du logiciel.

Je vais répondre tout de suite à des questions qu'on m'a déjà posées. L'objectif n'est pas de passer automatiquement les ordres. C'est facilement envisageable avec les intermédiaires étrangers (les français n'aiment pas trop ça) mais ce serait dommage de se priver du cerveau humain comme formidable outil de décision et d'appréhension. De plus ne se positionnant pas sur de l'hyper-trading, on pourra sans problème réfléchir très très lentement. Par conséquent l'argent géré par le logiciel est virtuel, du virtuel indexé sur le portefeuille réel qu'il vous appartient d'ouvrir chez un broker et de gérer selon les saintes suggestions du logiciel.



Plus précisément et plus techniquement, le programme est capable :

- d'interfacer R avec le code python du squelette. R est le langage le plus puissant en matière de traitement statistique. En outre il est extrêmement utilisé, notamment en finance, et il devient donc possible de simplement plugger ce que les meilleurs scientifiques R&D proposent en matière d'analyse et d'optimisation boursière (au passage on rejoint encore une des étapes de la partie 1, la 6 : communauté).
- de communiquer de manière transparente entre ses modules et avec l'utilisateur, en local comme en réseau, de manière totalement transparente. Cela repose sur un système de message très simple d'utilisation qui permet d'étendre le système par des sous-systèmes indépendants (tant qu'ils respectent le protocole de communication)
- d'archiver les études et les données dans une base MySQL pour permettre de capitaliser les résultats et ainsi mener des recherches sur de larges échelles.
- de proposer une analyse complète des performances des trades, avec l'aide d'un puissant générateur graphique et d'une interface web dédiée aux expériences.
- d'envoyer des notifications sur les smartphones lorsque des suggestions sont disponibles ou que de lourds calculs sont terminés.
- d'optimiser génétiquement les algorithmes écrits par l'utilisateur.
- d'être commandé vocalement et de répondre (bon... petit goodie datant des heures "creuses" d'Alten...)

En outre tout une librairie quantitative est disponible, de même qu'un certains nombres d'algorithmes prédictifs, d'analyse de tendance sur données déstructurées.

Et enfin de nombreuses ressources amassées par votre serviteur qui connaissait qu'il y a 4 mois : livres, notes, liens, tout ce qu'il faut pour devenir un ninja de tous les domaines cités, les projets open-source à intégrer d'urgence dans le programme, une liste indécente d'idées et de concepts à explorer pour littéralement hacker les marchés financiers !



Conclusion

La première version de QuanTrade est déjà bidouillée par des contributeurs, les tests seront bientôt finalisés, et la version live sera mise en route à 9h30 à l'ouverture du marché. Autrement dit le plus pénible est terminé et il est temps d'utiliser tout ce code pour aller chercher des pépètes et des sujets de recherche géniaux.

A présent, parallèlement à la poursuite du développement, la communication auprès des communautés open-source et étudiantes va commencer, de même que la formalisation du business model (qui transpire d'un peu partout de ce document) et de fait la recherche de financement.

A en juger par les articles dans Forbes ou dans le New-York Times, au succès fulgurant de quantopian, à la qualité des ressources en connaissances et en projets libres, il y a un énorme pari à faire, passionnant à mener.

Il ne reste plus qu'à répondre à mon mail pour récupérer tes accès au tableau de bord, et à venir boire une bière cette semaine pour passer aux choses sérieuses !

