

Построение торговой стратегии на основе моделирования действий RL (Reinforcement learning) агента, с учетом характеристик графа корреляций между 136 российскими акциями

Глухов Никита, gevnik08@gmail.com

September 2020

Полная статистика результата: нажмите на меня ([ссылка на Яндекс.Диск](#)).

1 Описание использовавшихся источников и данных

Моя работа, в основном, построена на своих соображениях и на своем анализе данных. Однако, в первоначальную версию работы легли идеи из этой статьи [статьи](#). Главная идея данной статьи была построение стратегии, которая динамично выбирает акции в портфель, в зависимости от их характеристик в графе корреляций.

Сеть акций строилась на основе корреляций log-доходностей акций, с окном dt . В статье использовалась линейная корреляция Пирсона.

Далее, матрица корреляций Китайских акций была переведена в матрицу расстояний, по формуле $d(i, j) = \sqrt{2(1 - \rho(i, j))}$. Матрица расстояний переводится в взвешенный граф акций, где вес ребра между вершинами i, j равен $d(i, j)$. После этого был применен алгоритм MST, который помогает убрать лишние соединения между вершинами графа (вершины = акции). Подразумевается, что мы оставляет только те ребра, которые позволяют нам пройти от одной вершины до другой с минимальным количеством "издержек" перехода, в данном случае издержки - расстояние между вершинами. То есть, чем больше издержек, тем больше расстояния, а чем больше расстояния, тем меньше корреляция между активами.

Далее, в статье выбирались два типа портфелей, центральные и периферийные. Принцип выбора портфелей был основан на характеристиках вершин. Так, например, центральные акции - это те акции, которые являются ло-

кальным "центром"тяжести корреляций для других акций. Переферийные акции - это те, которые обладают меньшими значениями "центральности". Здесь хорошо подойдет идея сравнения домов в центре Москвы, например, на краю Москвы.

В статье было обнаружено, что в "хорошие"дни для рынка (определяются индикаторами моментумов), следует выбирать $p\%$ из центральных акций, ибо они приносят лучше доходность, однако в ситуации "плохих"дней для рынка, стоит выбирать переферийные. По некоторыми своим причинам¹, я решил отойти от идеи репликации данной статьи и скорее погрузиться в хороший анализ данных основанный на идеях из этой статьи и собственных соображений

Данные были скачены через API Тинькофф Инвестиции. Данные дневные (OHLCV) на 136 российских акций начиная с 2010 года, 1 января, заканчивая 17 сентября 2020. В основном работа проводилась с ретернами, поэтому, где ретерны отсутствовали(NAN) по каким-то причинам (отсутствие торгов, например), они заменялись на 0. Таким образом, данные ретернов непрерывны. Цены не подвергались изменениям.

Для торговой стратегии используется одна акция с тикером 'VSMO'. Причина выбора данной акции проста - эта акция показывала стабильный рост почти все время существования. Если алгоритм способен хотя бы на таком рынке показывать профит, то уже хорошо. В ином случае пока что не имеет смысла рассматривать другие акции.²Разумеется, что при других ситуациях (более волатильный рынок, например) алгоритм может показать себя лучше, однако для упрощения задачи мы пока откладываем другие акции.

2 Краткое описание идеи

Наш анализ будет разделен на три составляющих:

1. Анализ доходностей акции
2. Анализ характеристики сети акций, построенной на основе корреляций ретернов с остальными акциями из выборки
3. Комбинирование переменных (фичей), найденных в 1) и 2) и поиск "подходящих которые мы определим как переменные значимо определяющие доходность $\log P_t - \log P_{t-1}$

В глобальном контексте, наша задача найти переменные (далее - фичи), которые помогут RL агенту оптимально выбирать действия short, hold или long каждый торговый день. (мы определяем оптимальность как выбор действий a_1, a_2, \dots, a_t , максимизирующие Net Asset Value в конце тестового

¹Одна из таких - мне очень хотелось попробовать RL в трейдинге, а с первоначальной идеей статьи это превращалось в экспоненциально сложную затею

²СПОЙЛЕР: он не показал хорошие результаты

периода. NAV строится на основе суммы дневным доходностей за период. Т.е. $NAV_0 = 1$, далее если $R_1 = -0.5, R_2 = 0, \dots, R_{t-1} = 0, R_t = +0.6$, то $NAV_t = 1 - 0.5 + 0.6 = 1.1$.³

Поэтому, одной из фичей будет оценка \hat{R}_t нашего $R_t = \log P_t - \log P_{t-1}$, также базовой фичей, которая скорее выступает как техническая (для работы алгоритмов RL) - сама доходность R_{t-1}

Далее, мы рассмотрим два "вида" доходностей: одна стандартная - это обычная $R_t = \log P_t - \log P_{t-1}$, другая будет сумма ΣR_t за окно *window*. Последняя "доходность" примерно аналогична взятию доходностей $R_t = \log P_t - \log P_{t-window}$. Последняя, по моей гипотезе, может хорошо (значимая предсказательная способность) описывать "глобальный" (заметный на более больших промежутках времени) моментум благодаря своим mean reversion характеристикам). Мы фильтруем ряд ретернов с помощью Python библиотеки ARCH. Мы используем стандартную constant mean - GARCH модель для филттрации данного ряда. Остатки этой модели мы будем оценивать и переводить в оцененные ретерны. Оценка остатков будет основываться на также отфильтрованном ряде ΣR_t , с ARX-EGARCH (non-constant mean) и на свойства сети корреляций.

[Гипотеза о корреляциях] Сильно обобщая, скажем, что у акции есть три состояния на промежутке времени больше 1 дня - она явно растет, ходит бочком или явно падает. Акция может явно падать или расти по двум основным причинам: а) собственная причина (что-то связанное конкретно с компанией, или конкретно с рынком этой акции, спекулянты) и/или б) причина глобальная (кризис, пятница, настроения). Если рост происходит по последней причине, то мы должны будем заметить изменения в корреляциях с другими акциями.

Описав корреляции как граф, а корреляции акций как вершины этого графа, то взяв характеристики вершин мы можем проследить глобальные изменения в корреляциях этой компании с другими. Характеристики считаются с помощью Python библиотеки networkx. Например, на Рисунке 1 мы можем наблюдать одну из характеристик центральности акции в сети, посчитанной скользящим окном в 20 дней.

Объединяя отфильтрованный ΣR_t с тремя параметрами, полученными анализируя сеть акций, мы используем данный набор данных вместе с библиотекой XGBoost для построения прогноза. Выбор данного метода оценки в основном базируется на свойстве деревьев - это feature importance, то есть таким образом мы сможем проверить эффективность наших переменных в плане предсказательной способности.⁴

³В таком формате было удобно построить первоначальную среду для обучения RL-агента, однако в такой среде есть недостатки. Обо всех таких и прочих недостатках будет сказано в конце

⁴Разумеется, можно было использовать обычный Random Forest, который менее подвержен оверфиттингу, но мне очень хотелось поработать с xgboost. К тому же на тестовой выборке он показал довольно хорошее предсказание данных

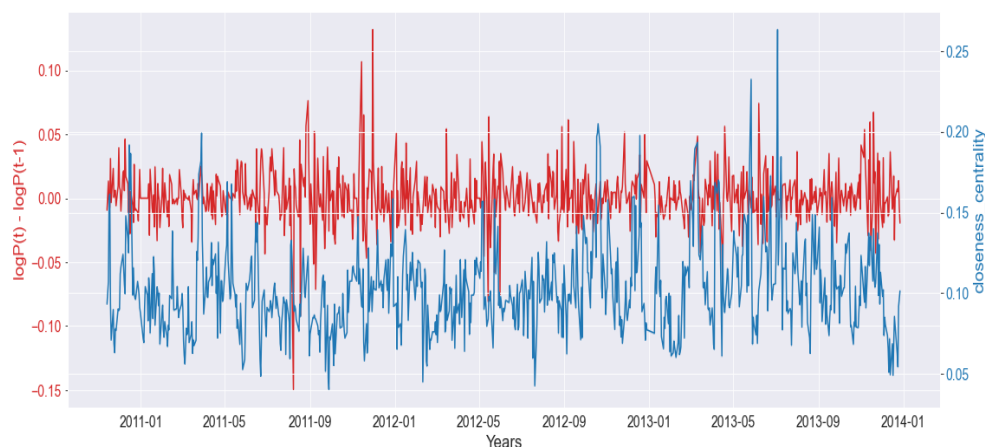


Рис. 1: График доходностей (красный) и меры центральности акции, посчитанная скользящим окном в 20 дней.

Полученные оценки и прежние фичи мы используем как состояние среды для RL агента.

Агент обучается с помощью алгоритма Deep Q-learning. Внутри применяется обычная MLP нейросеть написанная с помощью API Tensorflow. Для принятия решения применяется epsilon-greedy policy, с $\epsilon = 0.5$ и epsilon-reduction = 0.99.

3 Краткое описание результатов

Результативности стратегии, определенной как стабильный и положительный, доход на протяжении всего периода, достичь не удалось **не удалось**. Графики, показывающие основную результативность стратегии, показаны на Рисунке 2.

Полное количественное описание результата доступно по [ссылке](#).

Однако, в результате работы и анализа в целом, были выявлены интересные закономерности. Например, волатильность предсказанных с помощью наших регрессоров ретернов, имеют интересную зависимость с actual волатильностью ретернов (см. Рисунок 3). Интересно, что можно заметить, что волатильность предсказанных ретернов 'спешит', по сравнению с волатильностью actual ретернов. То есть, мы можем наблюдать увеличения в волатильности предсказанных ретернов, пока в настоящих увеличения нет. Однако, заметно, что, в среднем, волатильность настоящих доходностей "догоняет". Это интересное наблюдение, которое можно использовать, например, в предсказаниях волатильности. Или, в целом, таймингах, когда локальная структура волатильности (её среднее, максимальное значение на промежутке и тд) может меняться. Все это в дальнейшем можно исследовать.

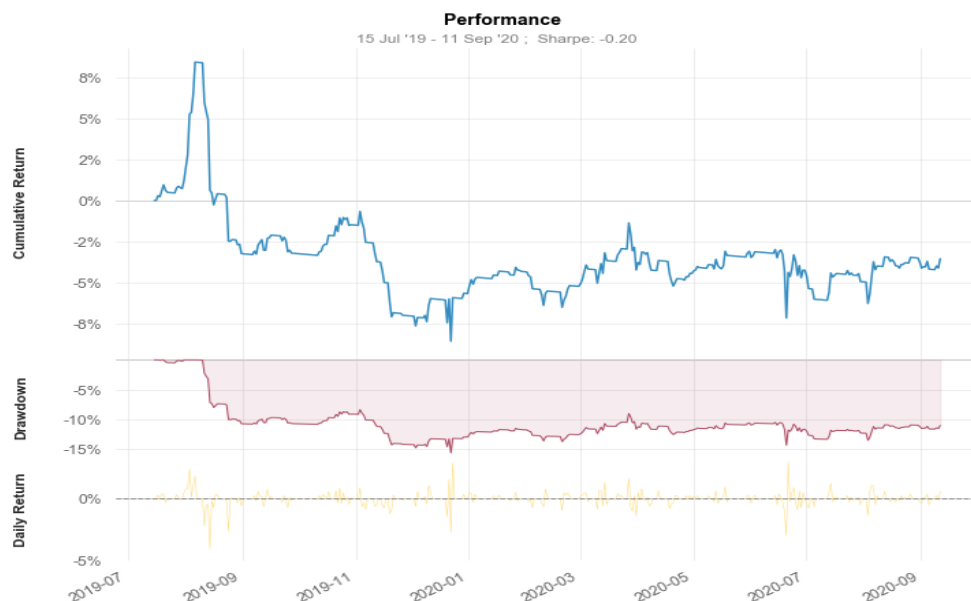


Рис. 2: График доходностей (красный) и меры центральности акции, посчитанная скользящим окном в 20 дней.

довать в контексте прайсинга и трейдинга опционов.

В целом, что было замечано в работе, так это то, что предсказанные ретерны действительно помогают хорошо определять тайминги волатильности. Однако, возникают проблемы предсказания *направления* ретернов. Наша модель понимает примерно когда будут происходить структурные изменения в волатильности, однако плохо понимает в какую сторону будут изменения (вверх, вниз или бочком). На текущий момент, это моя основная гипотеза почему уже агент не смог адекватно обучиться. Решением данной проблемы я нахожу в добавлении другого индикатора импульса. Так, мы гипотетически сможем не только таймить волатильность, но и пытаться "угадывать" направления изменений.

4 Риск Менеджмент

VaR был посчитан стандартным способом, как было показано на семинаре по риск менеджменту в одномерном случае.

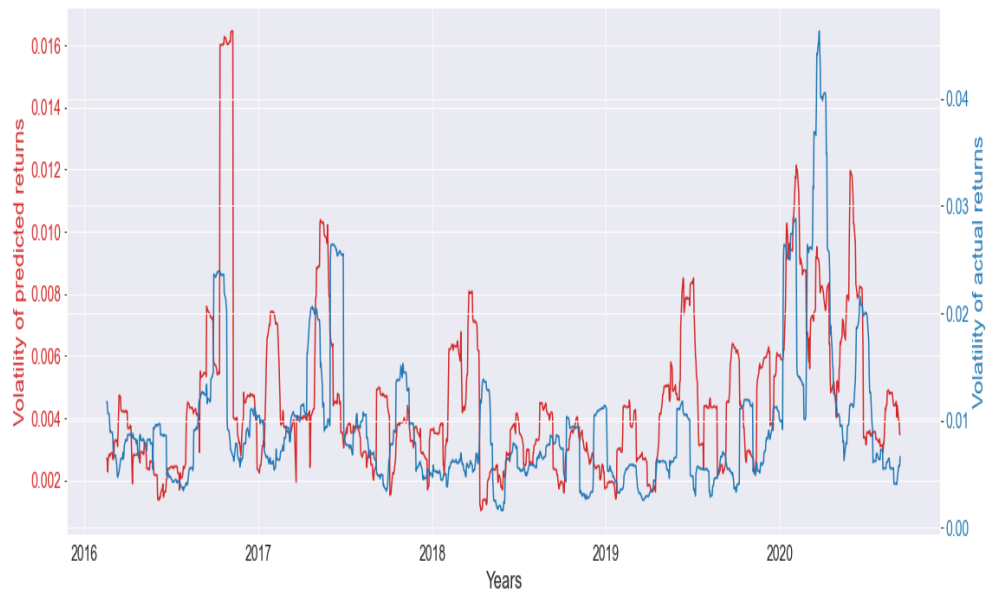


Рис. 3: Волатильность предсказанных ретернов и волатильность actual ретернов на том же промежутке времени.

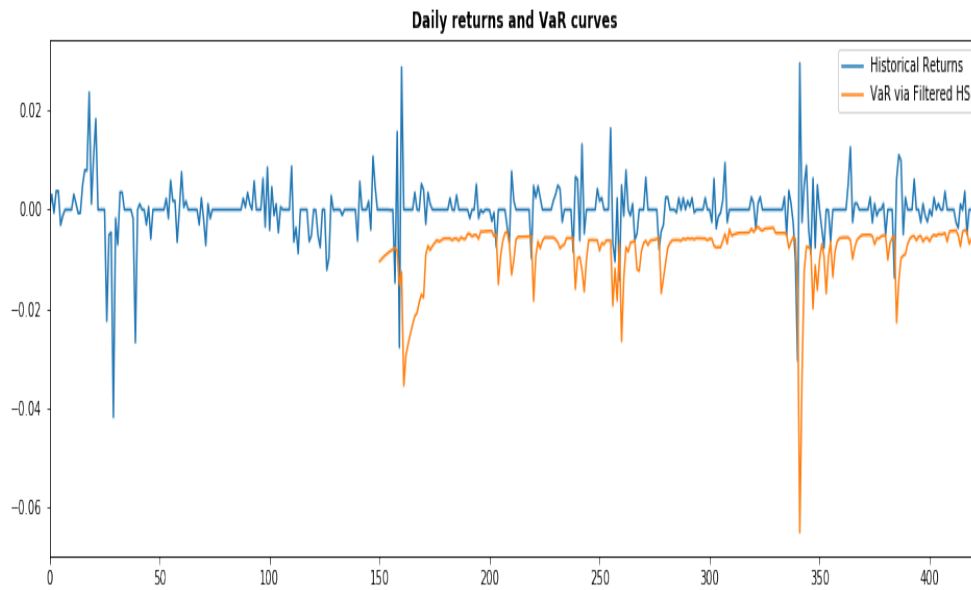


Рис. 4: Волатильность предсказанных ретернов и волатильность actual ретернов на том же промежутке времени.